

**ANÁLISE ESPACIAL DA EVOLUÇÃO DA
DISTRIBUIÇÃO DA POPULAÇÃO NA PENÍNSULA
IBÉRICA (1877-2001)**

Rui António Marques Gouveia

Trabalho de Projeto apresentada(o) como requisito parcial
para obtenção do grau de Mestre em Estatística e Gestão de
Informação

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

ANÁLISE ESPACIAL DA EVOLUÇÃO DA DISTRIBUIÇÃO DA POPULAÇÃO NA PENÍNSULA IBÉRICA (1877-2001)

por

Rui António Marques Gouveia

Trabalho de Projeto apresentada(o) como requisito parcial para a obtenção do grau de Mestre em Estatística e Gestão de Informação, Especialização em Análise e Gestão de Informação

Orientadora: Professora Doutora Ana Cristina Costa

Coorientador: Professor Doutor Daniel Alves

Novembro de 2015

DEDICATÓRIA

À minha família e aos meus Pais. Em particular ao meu Pai que nos deixou no decorrer deste Mestrado e, tal como em outras ocasiões semelhantes, tão entusiasmado estava com a minha frequência académica.

AGRADECIMENTOS

A realização deste de Mestrado só foi possível graças à colaboração e ao contributo, de forma directa ou indirecta, de várias pessoas, às quais gostaria de exprimir algumas palavras de agradecimento e profundo reconhecimento, em particular:

à Professora Doutora Cristina Costa, pela disponibilidade manifestada para orientar este trabalho, pela preciosa ajuda na definição do objeto de estudo, pela exigência de método e rigor, pela incansável orientação científica, pela revisão crítica do texto, pelos profícuos comentários, esclarecimentos, opiniões e sugestões, pela cedência e indicação de bibliografia relevante para a temática em análise, pelos oportunos conselhos, acessibilidade, cordialidade e simpatia demonstradas, pela confiança que sempre depositou em mim e pelo permanente estímulo que, por vezes, se tornaram decisivos em determinados momentos da elaboração deste trabalho de projeto, pelo interesse evidenciado, incluindo o benéfico acompanhamento ao longo do meu percurso académico, como Professora das disciplinas de Estatística durante o Mestrado no âmbito do qual é realizado este trabalho de projeto;

ao Professor Doutor Daniel Alves pela disponibilidade manifestada para coorientar este trabalho, pela exigência de método e rigor, a incansável orientação científica, revisão crítica do texto, pelos profícuos comentários, esclarecimentos, opiniões e sugestões, pela cedência e indicação de bibliografia relevante para a temática em análise, pelos oportunos conselhos, acessibilidade, cordialidade e simpatia demonstradas e pela confiança que sempre depositou em mim;

à Mestra Carolina Fonseca pela revisão do texto sob o ponto de vista gramatical e ortográfico, pontuação e adaptação do texto ao Novo Acordo Ortográfico em vigor.

por último, mas não menos importante, à minha esposa e aos meus pais, pelo apoio e compreensão inestimáveis, pelos diversos sacrifícios suportados e pelo constante encorajamento a fim de prosseguir a elaboração deste trabalho.

A todos, enfim, reitero o meu apreço e a minha eterna gratidão.

RESUMO

Ao longo da história os factores geográficos como altitude, precipitação, temperatura, distância aos rios, distância ao litoral explicaram sempre grande parte da distribuição espacial da população. Esta distribuição pode ser observada por variáveis como a densidade populacional de cada área geográfica e outras dela decorrentes.

Com a industrialização, maioritariamente desenvolvida nos maiores centros urbanos, com a diminuição dos custos de transporte e com a passagem de uma agricultura trabalho-intensiva a uma que faz uso de maquinaria, estes factores geográficos deixam de condicionar a distribuição espacial da população na Península Ibérica, à semelhança do que aconteceu noutros países.

Procurando aprofundar anteriores trabalhos exploratórios com metodologias de análise de *clusters* e regressão espacial, este trabalho de projecto visa analisar a distribuição geográfica da população na Península Ibérica ao nível administrativo da freguesia utilizando dados dos Censos recolhidos no período 1878-2001. Em particular, pretende-se investigar a hipótese de investigação de que os factores geográficos explicam grande parte da distribuição espacial da população da Península Ibérica até à II Guerra Mundial (1878-1940), e que no período 1940-2001 esses factores perdem importância sendo substituídos por factores socioeconómicos que conduzem à atracção das grandes cidades.

Os resultados obtidos através das análises de regressão para os dois períodos, trazem novas evidências para suportar esta hipótese. O número de segmentos (*clusters*) que optimiza o espaço ibérico é de quatro nos dois períodos. Em ambos os casos são bastante heterogéneos e atravessam a fronteira entre os dois estados ibéricos. Há mudanças de *clusters* entre as duas análises originadas maioritariamente pelas mudanças demográficas entretanto ocorridas. As análises de regressão espacial mostram que se verificou uma perda de influência das variáveis geográficas na distribuição da população na Península Ibérica do primeiro para o segundo período.

PALAVRAS-CHAVE

Análise de Clusters; Distribuição Espacial da População; História; Sistemas de Informação Geográfica; Península Ibérica

ABSTRACT

Throughout history, geographic factors such as altitude, precipitation, temperature, distance from rivers and coastline have always explained great part of the population's spatial distribution. This distribution can be observed by variables, such as the population density of each geographical area and others emerging from it.

With industrialization, mostly developed in major urban centres, with the reduction of transport costs and with the transition from a labour-intensive agriculture to an agriculture using machinery, these geographic factors no longer condition the population's spatial distribution in the Iberian Peninsula, similarly to what happened in other countries.

Seeking to deepen previous exploratory works with methodologies of cluster analysis and spatial regression, this project seeks to analyse the geographical distribution of population in the Iberian Peninsula at an administrative level of the parish using data from the censuses from 1878-2001.

Particularly, we intend to investigate the possibility that geographic factors explain great part of the population's spatial distribution in the Iberian Peninsula up to the 2nd World War (1878-1940), and that between 1940 and 2001 those factors become less important by being replaced by socioeconomic factors leading to the attraction of large cities.

The results obtained through the regression analysis for both periods bring new evidence to support this assumption. The number of segments (clusters) optimizing the Iberian space is of four years in the two periods. In both cases, they are quite heterogeneous and trespass the border between the two Iberian states. There are clusters' changes between the two analysis mainly arising from the demographic changes occurred in the meanwhile. The spatial regression analysis show a loss of influence from geographic variables in the population distribution throughout the Iberian Peninsula, from the first to the second period.

KEYWORDS

Cluster analysis; Spatial distribution of population; History; Geographic information systems; Iberian Peninsula

ÍNDICE

1. Introdução	1
1.1. Contexto e Identificação do Problema	1
1.2. Objectivo de Estudo	2
2. Revisão da Literatura	3
3. Dados	5
3.1. Granularidade dos Dados	5
3.2. Variáveis de Análise	7
4. Metodologias	8
4.1. SEM – <i>Spatial Error Model</i>	10
4.2. SLM – <i>Spatial Lag Model</i>	11
4.3. GWR – <i>Geographic Weighted Regression</i>	11
5. Resultados e Discussão	13
5.1. Segmentação	13
5.1.1. Evolução da análise e resultados obtidos	14
5.1.2. Caracterização dos Segmentos	19
5.1.3. Comparação dos dois períodos	22
5.2. Regressão Espacial	26
5.2.1. Período 1878-1940	27
5.2.2. Período 1940-2001	37
5.2.3. Comparação entre os Dois Períodos	47
6. Conclusões	49
6.1. Limitações e recomendações para trabalhos futuros	50
7. Anexos	51
8. Bibliografia	59

ÍNDICE DE FIGURAS

Figura 1 - Esquema de relacionamento do Modelo de Erro Espacial	10
Figura 2 - Esquema de relacionamento do Modelo de Desfasamento Espacial.....	11
Figura 3 - Distribuição dos valores de Curvatura média do polígono.....	13
Figura 4 - Mapa dos <i>clusters</i> obtidos sem os dados referentes à província de Madrid no período 1878-1940.....	16
Figura 5 - Mapa dos <i>clusters</i> obtidos com os dados referentes à província de Madrid no período 1878-1940.....	17
Figura 6 - Mapa dos <i>clusters</i> obtidos sem os dados referentes à província de Madrid no período 1940-2001.....	18
Figura 7 - Mapa dos <i>clusters</i> obtidos com os dados referentes á província de Madrid no período 1940-2001.....	19
Figura 8 - Gráfico de Coordenadas Paralelas de cada Variável dentro de cada <i>cluster</i> e número de polígonos desse <i>cluster</i> relativos ao período 1878-1940.....	20
Figura 9 - Gráfico de Coordenadas Paralelas de cada Variável dentro de cada <i>cluster</i> e número de polígonos desse <i>cluster</i> relativos ao período 1940-2001.....	20
Figura 10 - Mapa de decis da Taxa de Crescimento Anual Médio 1940-2001.....	21
Figura 11 - Mapa das mudanças de <i>cluster</i> entre 1878-1940 e 1940-2001	23
Figura 12 - Coeficientes da variável Altitude Média por polígono da regressão GWRT1A1S1R0L0P1T0CL1.....	29
Figura 13 - Coeficientes da variável Declive Médio por polígono da regressão GWRT1A1S1R0L0P1T0CL1.....	30
Figura 14 - Coeficientes da variável Precipitação Média por polígono da regressão GWRT1A1S1R0L0P1T0CL1.....	30
Figura 15 - Resíduos da regressão GWRT1A1S1R0L0P1T0CL1	31
Figura 16 - Coeficientes da variável Altitude Média por polígono da regressão GWRT1A1S1R0L0P1T0CL2.....	33
Figura 17 - Coeficientes da variável Declive Médio por polígono da regressão GWRT1A1S1R0L0P1T0CL2.....	34
Figura 18 - Resíduos da regressão GWRT1A1S1R0L0P1T0CL2	35
Figura 19 - Coeficientes da variável Altitude Média por polígono da regressão GWRT2A1S1R0L0P1T0CL1.....	39
Figura 20 - Coeficientes da variável Declive Médio por polígono da regressão GWRT2A1S1R0L0P1T0CL1.....	39

Figura 21 - Coeficientes da variável Precipitação Média por polígono da regressão GWRT2A1S1R0LOP1T0CL1.....	40
Figura 22 - Resíduos da regressão GWRT2A1S1R0LOP1T0CL1	41
Figura 23 - Coeficientes da variável Altitude Média por polígono da regressão GWRT2A1S1R0LOP0T0CL2.....	43
Figura 24 - Coeficientes da variável Declive Médio por polígono da regressão GWRT2A1S1R0LOP0T0CL2.....	44
Figura 25 - Resíduos da regressão GWRT2A1S1R0LOP0T0CL2	45

ÍNDICE DE TABELAS

Tabela 1 - Área média de cada polígono por Província / Distrito em KM2.....	6
Tabela 2 - Variáveis de Análise	7
Tabela 3 - Codificação das Regressões	8
Tabela 4 - Taxas de Crescimento da Província de Madrid e restantes Regiões.....	15
Tabela 5 - Média de Altitude por Região	15
Tabela 6 - Quadro com Valores Médios de cada Variável dentro de cada <i>cluster</i> e número de polígonos desse <i>cluster</i> relativos ao período 1878-1940	19
Tabela 7 - Quadro com Valores Médios de cada Variável dentro de cada <i>cluster</i> e número de polígonos desse <i>cluster</i> relativos ao período 1940-2001	20
Tabela 8 - Quadro com Somatório de População por <i>Cluster</i> de origem e de chegada para os censos de 1878, 1940 e 2001	24
Tabela 9 - Matriz de transferências de polígonos em termos de alterações na TCAM entre os dois períodos	25
Tabela 10 - Densidade Populacional da Matriz de transferências de polígonos da Tabela anterior.....	25
Tabela 11 - Tabela comparativa do critério AIC com e sem polígonos isolados.....	26
Tabela 12 - Diagnóstico de Dependência Espacial da Regressão T1A1S1L0R0P1T1CL1OLS ...	27
Tabela 13 - Diagnóstico de Dependência Espacial da Regressão T1A1S1L1R1P1T1CL1OLS ...	27
Tabela 14 - Coeficientes e testes da Regressão T1A1S1L0R0P1T1CL1SEM	27
Tabela 15 - Coeficientes e testes da Regressão T1A0S1L0R0P1T1CL1SEM	28
Tabela 16 - Coeficientes e testes da Regressão T1A1S1L0R0P1T1CL1SLM	28
Tabela 17 - Diagnóstico de Dependência Espacial da Regressão T1A1S1L0R0P1T1CL2OLS ...	31
Tabela 18 - Diagnóstico de Dependência Espacial da Regressão T1A1S1L1R1P1T1CL2OLS ...	32
Tabela 19 - Diagnóstico de Dependência Espacial da Regressão T1A1S1L0R0P1T1CL3OLS.....	35
Tabela 20 - Diagnóstico de Dependência Espacial da Regressão T1A1S1L1R1P1T1CL3OLS ...	36
Tabela 21 - Diagnóstico de Dependência Espacial da Regressão T1A1S1L0R0P1T1CL4OLS ...	36
Tabela 22 - Coeficientes e Testes da Regressão T1A1S1L0R0P1T1CL4SEM	36
Tabela 23 - Coeficientes e Testes da Regressão T1A1S0L0R0P0T1CL4SEM	37
Tabela 24 - Diagnóstico de Dependência Espacial da Regressão T2A1S1L0R0P1T1CL1OLS ...	37
Tabela 25 - Coeficientes e Testes da Regressão T2A1S1L0R0P1T1CL1SEM	37
Tabela 26 - Coeficientes e Testes da Regressão T2A1S1L0R0P0T0CL1SEM	38

Tabela 27 - Diagnóstico de Dependência Espacial da Regressão T2A1S1L0R0P1T1CL2OLS ...	41
Tabela 28 - Diagnóstico de Dependência Espacial da Regressão T2A1S1L1R1P1T1CL2OLS ...	41
Tabela 29 - Coeficientes e Testes da Regressão T2A1S1L1R1P1T1CL2SEM	42
Tabela 30 - Coeficientes e Testes da Regressão T2A1S1L0R0P1T0CL2SEM	42
Tabela 31 - Diagnóstico de Dependência Espacial da Regressão T2A1S1L0R0P1T1CL3OLS ...	45
Tabela 32 - Diagnóstico de Dependência Espacial da Regressão T2A1S1L1R1P1T1CL3OLS ...	45
Tabela 33 - Coeficientes e Testes da Regressão T2A0S1L1R1P1T1CL3SEM	46
Tabela 34 - Coeficientes e Testes da Regressão T2A1S1L0R0P0T1CL3SLM.....	46
Tabela 35 - Diagnóstico de Dependência Espacial da Regressão T2A1S1L0R0P1T1CL4OLS ...	46
Tabela 36 - Coeficientes e Testes da Regressão T2A1S1L0R0P1T1CL4SLM.....	47
Tabela 37 - Coeficientes e Testes da Regressão T2A0S1L0R0P1T1CL4SLM.....	47
Tabela 38 - Quadro comparativo de AIC entre períodos	48
Tabela 39 - Lista de Polígonos sem Dados Climatéricos	51
Tabela 40 - Lista de Polígonos sem População.....	52
Tabela 41 - Lista de Polígonos Isolados após Segmentação 1878-1940	56

LISTA DE SIGLAS E ABREVIATURAS

ALT_MEAN	Altitude Média
DIST_LITOR	Distância ao Litoral
DIST_RIOSP	Distância aos Rios Principais
GWR	Geographic Weighted Regression
OLS	Ordinary Least Squares
SEM	Spatial Error Model
SLM	Spatial Lag Model
SLP_MEAN	Declive Médio
TCAM 1878-1940	Taxa de Crescimento Médio da População no Período 1878-1940
TCAM 1940-2001	Taxa de Crescimento Médio da População no Período 1940-2001
WCPRECMEAN	Precipitação Média
WCTEMPMEAN	Temperatura Média

1. INTRODUÇÃO

Os factores geográficos sempre influenciaram a distribuição espacial da população. Nos últimos dois séculos, com a industrialização e redução de custos de transporte, inicialmente associados à introdução dos caminhos-de-ferro (da Silveira, Alves, Lima, Alcantara, & Puig, 2011), ambos a ocorrerem em diferentes momentos do tempo nesta janela temporal nos diferentes países europeus, existiu um menor peso destes factores como explicação para a localização dos seres humanos.

Deixou de existir uma necessidade de proximidade do espaço rural para suprir a alimentação (Marti-Henneberg, 2005). Os ganhos crescentes de escala (Krugman, 1991), com a consequente possibilidade de melhores ganhos salariais (Rappaport & Sachs, 2003) levaram as pessoas a migrar para esses grandes centros urbanos.

1.1. CONTEXTO E IDENTIFICAÇÃO DO PROBLEMA

O número de estudos sobre a distribuição geográfica da população na Península Ibérica ao nível administrativo da freguesia em Portugal e do município em Espanha é extremamente limitado. Existem estudos para outras áreas geográficas sobre este tema mas com uma granularidade geográfica inferior que só vão ao nível dos distritos, cidades e concelhos.

Ayuda, Collantes, & Pinilla, (2010a) verificam que em Espanha os factores geográficos, que fundamentam a localização, através dos factores socioeconómicos – a distribuição dos sectores com lucros crescentes, são a principal explicação para a concentração da população espanhola desde o início do século XX. Pons, Paluzie, Silvestre, & Tirado, (2007) demonstram a existência de uma relação directa entre a decisão de localização dos trabalhadores e o potencial de mercado das regiões anfitriãs. Estes resultados são idênticos aos de Le Gallo & Chasco, (2008). Estes autores estudam os municípios espanhóis ao longo do século XX. Finalmente, Ayuda, Collantes, & Pinilla, (2010b) demonstram que estes movimentos migratórios se intensificam em Espanha durante a segunda metade do século XX.

Da Silveira, Alves, Painho, Costa, & Alcantara, (2013) já desenvolveram um trabalho sobre o espaço geográfico Península Ibérica e não só Espanha. Da Silveira et al., (2013) conseguem provar que a fronteira política, não é geográfica, o que só seria de supor a partir da adesão à então Comunidade Económica Europeia em 1986. Um aprofundamento do estudo já feito seria interessante para corroborar os resultados obtidos por estes autores mas também por outros.

Sabemos que há esta tendência de centralização em grandes áreas urbanas (Fujita & Mori, 1996) e litoralização (Rappaport & Sachs, 2003) mas é uma constatação empírica que precisa de ser comprovada com as novas metodologias de análise de informação e com *software* de informação geográfica, especificamente para este espaço geográfico no extremo sudoeste da Europa.

1.2. OBJECTIVO DE ESTUDO

O nosso objectivo de estudo é investigar a hipótese – até início da II Guerra Mundial (1877-1940) os fatores geográficos explicam grande parte da distribuição espacial da população; no período 1940-2001, foram os fatores socioeconómicos (atração das grandes cidades) que passaram a explicar as alterações nos padrões da população.

Deste objectivo principal derivam os seguintes objectivos específicos os quais caracterizam esta problemática:

- ❑ Qual o número de *clusters* que optimiza a distribuição do espaço ibérico?
- ❑ Quais as variáveis geográficas que explicam a densidade populacional na Península Ibérica?
- ❑ Qual a significância das variáveis: Altitude média do polígono, Declive médio do polígono, Precipitação média do polígono, Temperatura média do polígono, Altitude da sede do polígono, Distância da sede ao litoral, Distância da sede aos rios principais, na distribuição geográfica da população?
- ❑ Qual o modelo de regressão espacial que se adequa aos dados em análise?

2. REVISÃO DA LITERATURA

Não foi possível identificar literatura com uma abordagem metodológica igual à que se pretende aplicar, para qualquer área geográfica nos países industrializados, em qualquer período de tempo. Assim, procuram-se analisar estudos que tivessem como objetivo testar hipóteses semelhantes para diferentes áreas geográficas e diferentes períodos temporais. Para este efeito, a revisão da literatura tem perseguido três direções principais. Primeiramente, procuraram-se estudos que tivessem como foco Portugal, Espanha ou a Península Ibérica. Em segundo lugar, a revisão da literatura cobriu estudos sobre outras áreas geográficas, em que existissem abordagens metodológicas o mais semelhante possível àquela por nós escolhida. Finalmente, observaram-se estudos cuja análise pudesse complementar e reforçar a hipótese testada ou qualquer ideia enquadradora. Em qualquer dos casos, e à exceção de Da Silveira et al., (2013), um trabalho onde se baseia a análise que pretendemos fazer, a granularidade é sempre inferior à considerada neste trabalho.

Os autores Ayuda et al., (2010a), testaram duas hipóteses. A primeira engloba os fundamentos geográficos como sementes para a instalação de populações. A segunda, foca as vantagens dos rendimentos crescentes como fatores concentradores de população. Este estudo, foca-se essencialmente no território espanhol, sobre a população de 1787 e 2000, explicando como se distribuía a população pelo território espanhol antes da industrialização e como esta veio influenciar a própria distribuição da população. Abordaram duas perspetivas para explicar a concentração da população desigual. Por um lado, os fatores ambientais como o relevo, clima, entre outros, são a principal causa de padrões de distribuição, podendo ser tanto diretos (agricultura) como indiretos (custos de transporte). Por outro lado, “a nova geografia económica”, tem vindo a ter uma grande importância ao gerar retornos crescentes nas novas atividades económicas desenvolvidas complementarmente com o surgimento da industrialização.

Este trio de autores Ayuda et al., (2010b) faz, neste outro trabalho, uma análise das disparidades de longo prazo do crescimento regional da população na Europa Continental. Propõe uma equação de convergência para oito países da Europa Ocidental no período de 1850-2000. Os resultados provam que a divergência no crescimento económico é um padrão por toda a Europa. É também em Espanha que este processo de divergência demográfica é testado, focando no processo de concentração-dispersão em grande escala. Uma particularidade de grande interesse para o nosso trabalho é a tipificação do crescimento de longo prazo da população espanhola através da análise de *clusters*.

Marti-Henneberg, (2005) enfatiza a questão de que se tem vindo a estudar a distribuição da população europeia como um agregado global. No seu artigo, o autor apresenta evidências para apoiar a tese de que a população europeia tem sofrido uma progressiva concentração regional, exacerbando as desigualdades regionais ao longo do período histórico estudado – 1870-2000. O seu principal contributo é estudar este fenómeno a partir de uma perspetiva histórica.

Fujita & Mori, (1996) analisam o crescimento das cidades portuárias nos Estados Unidos da América à luz da teoria comercial de ganhos crescentes e das políticas de transportes. Os ganhos crescentes existem quer ao nível dos bens de consumo, quer ao nível de fatores de produção intermédios. Este estudo exploratório é importante para corroborar a perda de significância dos fatores geográficos como variáveis explicativas para a distribuição das populações e a litoralização das cidades.

Rappaport & Sachs, (2003) fazem uma análise regressiva rica, rigorosa e exaustiva sobre as variáveis que influenciam a localização das populações ao longo das costas norte-americanas e das margens dos grandes lagos. Primeiramente, um efeito produtividade e, posteriormente, a qualidade de vida geram concentração costeira.

Pons et al., (2007) comprovam, com modelos da nova economia geográfica, a hipótese de que os trabalhadores migrantes escolhem cidades de destino com maior potencial de mercado e com maior dimensão. O padrão de migração altera-se com o começo da industrialização em várias províncias espanholas desde o fim do século XIX até à Guerra Civil espanhola.

Num artigo mais recente de Da Silveira et al., (2013) são apresentados alguns resultados exploratórios que indicam que, até ao final da II Guerra Mundial (1877-1940), os fatores geográficos explicavam grande parte da distribuição espacial da população e que já no período de 1940-2001, foram os fatores socioeconómicos (atração das grandes cidades) que passaram a explicar as alterações nos padrões da população. Este artigo estuda a Península Ibérica como um todo. Os autores realçam a sua intenção de verificar se a fronteira entre Espanha e Portugal tem alguma influência na distribuição da população pelo território. Este artigo é a base de toda esta proposta de tese, uma vez que será investigada a mesma hipótese, utilizando métodos descritivos e preditivos de análise.

3. DADOS

Com o objetivo de observarmos a influência dos fatores geográficos na distribuição da população, procurámos utilizar o maior nível de granularidade possível a nível administrativo – as freguesias portuguesas e os municípios espanhóis (designados polígonos). Os dados da população são oriundos dos 13 censos executados desde 1877 (Espanha) e 1878 (Portugal) até 2001. Como este trabalho é realizado na sequência do trabalho de Da Silveira et al., (2013), e ao utilizar os dados subjacentes a este trabalho exploratório, vamos utilizar a série homogénea da população residente no período 1900-2001 em Espanha (Ros, Rodríguez, Gisbert, & Mas, 2006), completada para o período 1877-1900, e com os dados reunidos para Portugal relativamente ao mesmo período.

As variáveis geográficas escolhidas foram a altitude, o declive, a variabilidade do declive, a curvatura do terreno, a precipitação anual média e a temperatura anual média. Levámos em conta a altitude média de cada freguesia portuguesa e de cada município espanhol. Não levámos em conta a altitude do aglomerado principal dado que este pode ter só uma pequena parte da população residente nesse polígono. Estes dados foram obtidos a partir de um modelo de terreno digital para toda a Península Ibérica com uma resolução espacial de 200m * 200m, construída para um projeto de investigação anterior de Da Silveira et al., (2013). Estes dados foram fornecidos pela Carta Administrativa Oficial de Portugal 2012 e, no caso espanhol, pelo Nomenclátor Geográfico de Municipios y Entidades de Población.

As variáveis precipitação média e temperatura média derivam de dados retirados dos *rasters* “ESRI grids” do projeto WorldClim (<http://www.worldclim.org/current>). Estes dados estavam disponíveis numa base mensal e são representativos do período 1950-2000. Foi efetuado o cálculo para valores anuais (temperatura anual média e total anual de precipitação); em seguida, foi utilizada a “Zonal Statistics as Table”, do *software* ArcMap, para extrair os valores das duas variáveis.

Contudo, o cálculo da “Zonal Statistics as Table” não gerou valores para 67 polígonos (vide Tabela 39). Uma vez que 33 destes polígonos se referem a freguesias com população, e a maior parte delas se situa em centros urbanos, optámos por fazer uma interpolação de dados simples com os valores anteriormente obtidos para os polígonos contíguos de 1.ª ordem. Os restantes 34 polígonos, que integram um grupo maior de 142, foram retirados da análise pelos motivos que abaixo se explicam.

Decidimos retirar 142 polígonos (listados na Tabela 40 dos Anexos) da análise. Durante todo o horizonte temporal nunca tiveram qualquer habitante e estão todos localizados em território espanhol. Em nomenclatura administrativa, 70 destes polígonos têm o nome sugestivo de território histórico.

A divisão das variáveis demográficas em dois períodos deve-se à aceleração do processo de industrialização dos países ibéricos a partir das décadas de 30 / 40. Em Portugal os planos de fomento do Estado Novo iniciados na década de 30 não são interrompidos por uma Guerra Civil como acontece em Espanha.

3.1. GRANULARIDADE DOS DADOS

De forma a obter uma análise mais pormenorizada e rigorosa das mudanças da distribuição da população ao longo das décadas em estudo, optámos pela granularidade utilizada no trabalho de Da

Silveira et al., (2013). Abaixo do município espanhol não há um nível inferior de granularidade. Utilizando o critério da área (em km²), podemos observar na Tabela 1 que o valor médio dos municípios por província espanhola é superior às freguesias por distrito português. Mas, apesar de a média global de Espanha ser 61,27 KM², e de a de Portugal ser de 27,06 KM², há distritos portugueses com valores médios superiores às províncias espanholas e vice-versa.

Tabela 1 - Área média de cada polígono por Província / Distrito em KM²

Província / Distrito	Média de Área em KM ² de cada Polígono
Múrcia	253,5403
Ciudad Real	194,9194
Córdoba	183,9911
Albacete	170,7999
Cádiz	169,1820
Lugo	147,1630
Jaén	138,2988
Asturias	136,0503
Sevilla	133,8457
Badajoz	132,8572
Huelva	128,2457
Beja	104,7552
Cáceres	90,7834
Almería	86,5935
A Coruña	84,5799
Évora	81,2298
Ourense	79,0675
Huesca	77,8715
Granada	75,6242
Toledo	75,5622
Málaga	73,2314
León	72,5364
Cuenca	72,4733
Pontevedra	72,4313
Portalegre	70,7406
Setúbal	62,9612
Teruel	62,8456
Total Geral Espanha	61,2750
Zaragoza	59,6532
Faro	59,4659
Álava	57,5660
Soria	56,3106
Lleida	53,2148
Cantabria	51,8002

Província / Distrito (cont.)	Média de Área em KM ² de cada Polígono
Castellón	49,6892
Madrid	44,7544
Zamora	42,6215
Guadalajara	42,4885
Palencia	42,2432
Alicante	41,7002
Castelo Branco	41,4158
Valencia	41,1774
Valladolid	36,1148
Santarém	35,1987
Tarragona	34,9193
Salamanca	34,0476
Burgos	33,7428
Segovia	33,0570
Ávila	32,5187
Navarra	31,0448
La Rioja	28,9877
Girona	27,2703
Barcelona	25,2876
Leiria	23,7105
Bragança	22,1431
Total Geral Portugal	22,0570
Guipózcua	21,8933
Vizcaya	20,0653
Coimbra	19,1065
Guarda	16,4764
Vila Real	16,1975
Viseu	13,4695
Aveiro	13,4591
Lisboa	12,6177
Viana	7,6531
Porto	6,0878
Braga	5,2645

Amarelo - Espanha

Laranja - Portugal

3.2. VARIÁVEIS DE ANÁLISE

As variáveis de análise são as seguintes:

Tabela 2 - Variáveis de Análise

Acrónimo	Designação da Variável
POPD1878	Densidade Populacional em 1878
POPR1940	População Residente em 1940
POPR2001	População Residente em 2001
TCAM1878-1940	Taxa de Crescimento Anual Média no Período 1878-1940
TCAM1940-2001	Taxa de Crescimento Anual Média no Período 1940-2001
ALT_MEAN	Altitude Média do Polígono
SLP_MEAN	Declive Médio do Polígono
DIST_LITORAL	Distância ao Litoral
DIST_RIOSP	Distância aos Rios Principais
WCPRECMEAN	Precipitação Média
WCTEMPMEAN	Temperatura Média

4. METODOLOGIAS

O trabalho desenvolveu-se em duas fases. Numa primeira etapa procedeu-se à segmentação dos polígonos através de uma análise de *clusters* pelo método *K-Means* (Han, 2006). Na segunda fase de análise, aplicaram-se diferentes metodologias de regressão espacial aos dados de cada cluster separadamente: *Ordinary Least Squares* (OLS), *Spatial Error Model* (SEM), *Spatial Lag Model* (SLM), *Geographic Weighted Regression* (GWR). Nas secções seguintes faz-se um breve resumo destes métodos, seguindo de perto a descrição apresentada no tutorial <http://www.s4.brown.edu/S4/courses/SO261-John/lab9.pdf>. Mais detalhes sobre os métodos de regressão espacial considerados podem ser consultados em Wooldridge, (2009), Ward & Skrede Gleditsch, (2008) e Brunson, Fotheringham, & Charlton, (1996).

O *software* utilizado para fazer a análise de *clusters* foi o JMP Pro 10. Foi efetuada uma normalização das variáveis em questão e o critério ao qual o estudo obedeceu foi o *Cubic Clustering Criterion* (CCC). É uma solução puramente “quantitativa”. Do ponto de vista da experiência de utilizadores profissionais, esta solução traduz-se, normalmente, num elevado número de *clusters*, o que torna complicada a sua interpretação e posterior operacionalização. Contudo, não foi o caso vertente.

A normalização dos dados foi efetuada no processo de segmentação para que as diferenças dos dados de todas as variáveis tivessem a mesma importância.

O *Software* utilizado para fazer as regressões dos mínimos quadrados ordinários (OLS), o modelo de erro espacial e o modelo de *lag* espacial foi o GeoDa. Exceto indicação em contrário, o nível de significância considerado nos testes estatísticos foi 5%.

Doravante, quando nos referirmos às várias regressões utilizaremos a nomenclatura constante da Tabela 3 para condensar a informação das variáveis, dos dados e do tipo de regressão feita. No primeiro período a variável dependente foi sempre a TCAM1878-1940, e no segundo período foi a TCAM1940-2001.

Tabela 3 - Codificação das Regressões

Código	Variável / Regressão	Definição
T1	Periodo 1878-1940	0 - Não utilizada 1 - Utilizada
T2	Periodo 1940-2001	
A	Altitude Média	
S	Declive Médio	
L	Distância ao Litoral	
R	Distância aos Rios Principais	
T	Temperatura Média	
P	Precepitação Média	
CL	Cluster	Número do Cluster
OLS	Regressão dos Minimos Quadrados Ordinários	Tipo de Regressão
SEM	<i>Spatial Error Model</i>	
SLM	<i>Spatial Lag Model</i>	

Tendo a noção de estarmos a lidar com dados geográficos, utilizámos a Regressão *Ordinary Least Squares* (OLS) para:

1. testar a multicolinearidade dos dados (número de condição de multicolinearidade);
2. analisar a normalidade dos resíduos (teste de Jarque-Bera);
3. analisar a heteroscedasticidade (Breusch-Pagan, Koenker-Bassett, and White);
4. fazer um diagnóstico de dependência espacial.

Primeiramente, deverá considerar-se o número de condição de multicolinearidade. Não é um teste estatístico *per se* mas um diagnóstico que pode sugerir problemas com a multicolinearidade das variáveis escolhidas. Um valor acima de 30 é sugestivo de problemas.

O teste de Jarque-Bera sobre a normalidade dos erros é uma estatística Qui-quadrado com 2 graus de liberdade. A obtenção de resultados significativos neste teste é condição necessária para procedermos à inferência estatística.

O próximo diagnóstico é um teste estatístico para detectar a heteroscedasticidade, isto é, uma variância não constante dos resíduos. É o teste de Breusch-Pagan.

A correlação espacial está quase sempre presente nos dados geográficos. Trata-se de uma lei bem conhecida da Geografia, a Primeira Lei de Tobler (Miller, 2004): localizações próximas tendem a ter valores mais idênticos do que localizações afastadas. Assim sendo, estamos na presença de dados não independentes, o que nos impede de utilizar o Modelo Clássico de Regressão Linear. Um dos pressupostos da regressão é a independência das observações. Caso tal não se verifique, obtemos estimativas imprecisas dos coeficientes e o termo de erro contém dependências espaciais, i. e., informação significativa. Uma das medidas de correlação espacial mais conhecida é o I de Moran (Moran, 1950) que, na sua versão local (Anselin, 1995), já foi aplicada aos dados a utilizar neste trabalho por Da Silveira et al., (2013), com resultados conducentes à necessidade de aplicação de métodos de regressão espacial.

Prosseguimos para a aplicação do método correto de regressão espacial, escolhido pelos Testes Estatísticos AIC e de Multiplicadores de Lagrange (Anselin, Bera, Florax, & Yoon, 1996): o modelo de desfasamento espacial (ou *Spatial Lag Model*) e o modelo do erro espacial (ou *Spatial Error Model*).

A Tabela de Diagnóstico de Dependência Espacial das Regressões reporta o I de Moran (Moran, 1950). Valores elevados desta estatística indicam presença de autocorrelação espacial, o que invalida o uso da Regressão dos Mínimos Quadrados Ordinários.

São reportados cinco testes estatísticos de Multiplicadores de Lagrange no Diagnóstico de Dependência Espacial. Os dois primeiros (LM-Lag e Robust LM-Lag) dizem respeito ao modelo de desfasamento espacial como alternativa. Os dois seguintes (LM-Error e Robust LM-Error) referem-se ao modelo de erro espacial como a alternativa. As versões robustas só devem ser consideradas quando as versões *standard* são significativas.

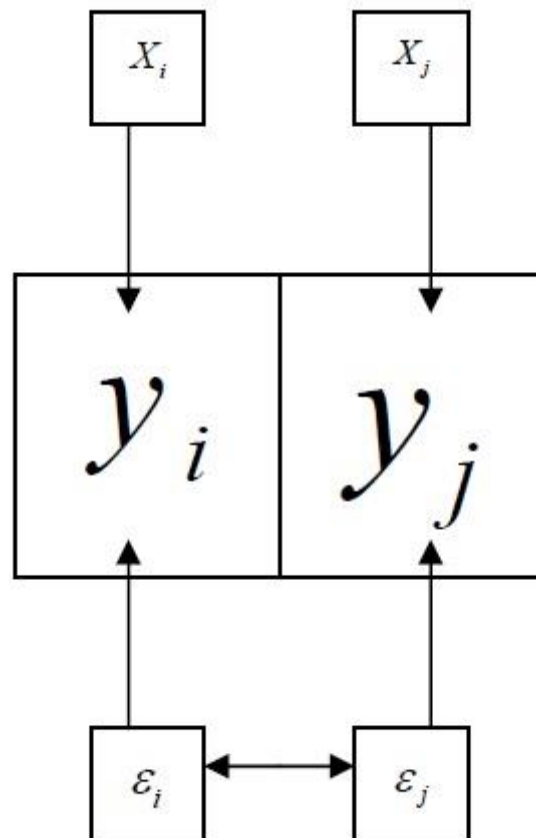
Se uma das estatísticas de teste LM rejeita a hipótese nula, e a outra não rejeita, então o prosseguimento será: estimar o modelo de regressão espacial alternativa que corresponde à estatística de teste que rejeita a hipótese nula.

Para calcular a multiplicidade de regressões feitas no âmbito deste trabalho utilizaram-se pesos de 2.^a contiguidade com esquina semelhantes aos do *paper* de Da Silveira et al., (2013). A escolha da 2.^a contiguidade pareceu-nos ser um bom compromisso entre a ausência de informação da vizinhança e um excesso de informação redundante. A escolha da contiguidade com esquina *versus* sem esquina resulta da nossa percepção de que as freguesias que se tocam só em esquinas são ainda próximas e se relacionam reciprocamente.

4.1. SEM – SPATIAL ERROR MODEL

No modelo *Spatial Error Model* (SEM) os erros das diferentes unidades espaciais estão correlacionados.

Figura 1 - Esquema de relacionamento do Modelo de Erro Espacial

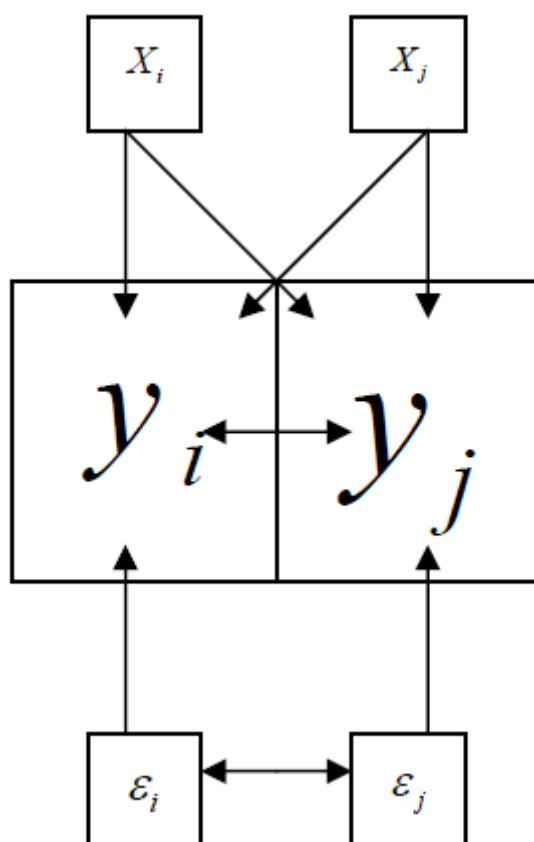


A existência de erro espacial na regressão OLS viola a assunção de erros não correlacionados. Como resultado, as estimativas são ineficientes. O modelo SEM procura lidar com esta questão é indicativo de covariações omitidas que, se não tidas em conta, afetam a inferência, como é patente na Figura 1.

4.2. SLM – SPATIAL LAG MODEL

Neste modelo a variável dependente y no lugar i é afectado pelas variáveis independentes nos lugares i e j . Com o Desfasamento Espacial na regressão OLS a assunção de erros não correlacionado é violada. Como resultado as estimativas são enviesadas e inefficientes. O Modelo de Desfasamento Temporal é sugestivo de um possível processo de difusão. Isto está esquematizado na Figura 2. Eventos num polígono prevêm uma crescente com uma Máxima Verosimilhança eventos nos polígonos vizinhos.

Figura 2 - Esquema de relacionamento do Modelo de Desfasamento Espacial



4.3. GWR – GEOGRAPHIC WEIGHTED REGRESSION

Face à qualidade de resultados patenteada nos modelos *Spatial Error Model* e *Spatial Lag Model*, fomos obrigados a recorrer ao modelo de Regressão Geográfica Ponderada. Por que não resolver o assunto da não estacionaridade de origem geográfica, permitindo que as relações entre as variáveis sejam diferentes no espaço em estudo? Esta é a essência do modelo GWR. Os resíduos deste modelo são genericamente mais baixos e não demonstram autocorrelação espacial, de acordo com os autores que propuseram este método Brunsdon et al., (1996).

O software utilizado para fazer as *Geographic Weighted Regressions* foi o ArcGIS 10.2.

Ao tentar ajustar os vários modelos de *Geographic Weighted Regression*, tivemos de tomar várias opções comuns a todos. Foram as seguintes:

1. Tipo de Kernel (Núcleo) – Especificação se o kernel é uma distância fixa ou se é adaptativo e tem uma distância variável em função da densidade. Optou-se por usar o núcleo adaptativo que permite a variação da distância e número de vizinhos consoante a densidade. O contexto espacial (o núcleo gaussiano) é uma função de um determinado número de vizinhos. Se a distribuição característica é densa, o contexto espacial é menor; onde a distribuição de recurso é escassa, o contexto espacial é maior;
2. Método de largura de banda - Especifica a forma como a extensão do núcleo deve ser determinada. Quando AICc ou CV são selecionados, a ferramenta irá encontrar o parâmetro à distância / vizinho ideal. Selecionou-se AICc - Akaike Information Criterion, dado que não se sabia qual o melhor valor a usar para os parâmetros da Distância (Tipo de Núcleo Fixo) ou do número de vizinhos (Tipo de Núcleo Adaptativo);
3. Pesos – O campo numérico que contém uma ponderação espacial para características individuais. Este campo de peso permite que algumas características sejam mais importantes no processo de calibração de modelo do que outras. É principalmente útil quando o número de amostras colhidas em diferentes locais varia. Os valores para as variáveis dependentes e independentes são a média e localizações com mais amostras são mais confiáveis (deve ser ponderada superior). Decidimos utilizar este ponderador.

Para definir os pesos referidos no parágrafo anterior foram escolhidos dois critérios, a saber:

1. Na conceptualização das relações espaciais escolhemos o Método de contiguidade espacial - Queen-Based Contiguity. Este método determina como unidades vizinhas, aquelas que têm qualquer ponto comum, incluindo fronteiras comuns e cantos comuns;
2. No método do cálculo da distância entre dois pontos foi escolhido o tradicional método da distância euclidiana.

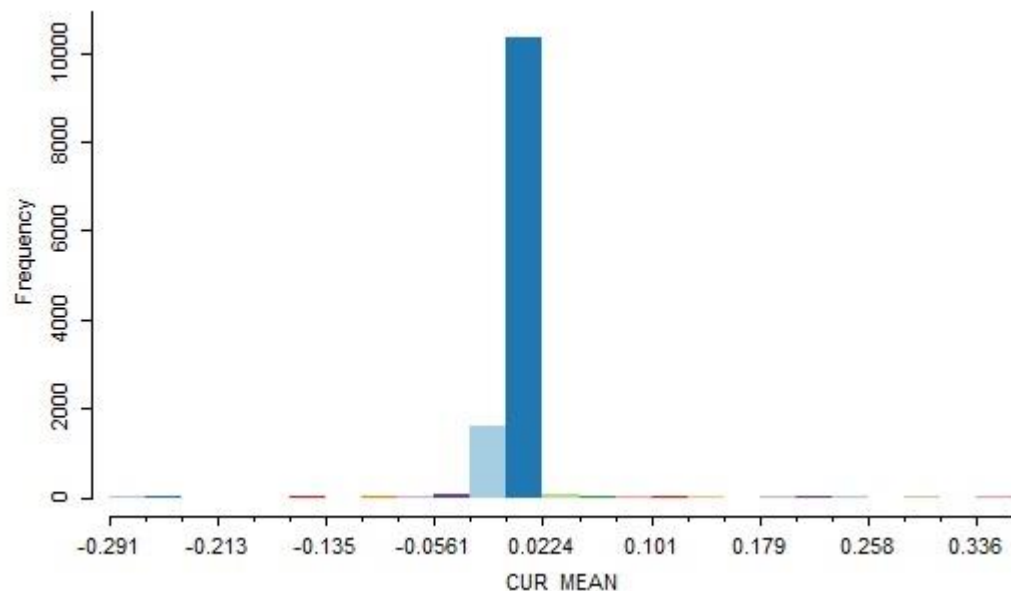
5. RESULTADOS E DISCUSSÃO

5.1. SEGMENTAÇÃO

Para verificar a hipótese em teste, procurámos dividir os dados nos dois períodos temporais em causa: 1878-1940 e 1940-2001. Se, por um lado, as variáveis geográficas são invariáveis à mudança de período, por outro lado, as variáveis demográficas não o são.

No decorrer dos trabalhos decidiu-se retirar a variável Curvatura. Esta variável descreve a concavidade / convexidade média do terreno em cada polígono: as vertentes retilíneas têm valor nulo de curvatura, as vertentes concavas têm valores positivos e as convexas têm curvatura negativa. É uma variável pouco informativa, cujos registos giram em torno do valor zero (Figura 3). A esmagadora maioria dos valores – mais de 11.000 (91%) – situa-se em duas classes.

Figura 3 - Distribuição dos valores de Curvatura média do polígono



Se queremos analisar dois períodos temporais, a taxa de crescimento médio anualizada (TCAM) dentro de cada um deles ajuda a perceber as semelhanças ao longo de cada período. Dentro de cada um dos intervalos temporais escolhidos acrescentou-se sempre a variável TCAM187840 ou TCAM194001, respetivamente. Para além dos dados geográficos e climatéricos – que são definidores de *clusters* –, estes dados demográficos também ajudam a encontrar semelhanças entre as várias unidades administrativas.

A escolha das restantes variáveis deveu-se aos seguintes fatores:

- Altitude – valores elevados de altitude significam, à partida, lugares inóspitos. Valores baixos significam proximidade de rios ou do litoral – zona para a qual há evidências, no período em estudo, para relocalizar a população como comprova Rappaport & Sachs, (2003) e Fujita & Mori, (1996);

- Declive – o declive médio do polígono também se traduz na hospitalidade / capacidade do terreno para permitir uma localização de um maior número de pessoas e de uma densidade populacional maior. Isto também pode ou não influenciar um crescimento tipo “mancha de óleo”. Ao nível económico também se prende com a viabilidade da agricultura. Terrenos com um determinado declive tornam-se inviáveis para a agricultura e esta foi, pelo menos durante boa parte do período estudado a principal atividade económica da população ibérica;
- Variabilidade do declive – o mesmo que a anterior;
- Precipitação média – a precipitação pode influenciar a localização positiva e negativamente: positivamente devido à existência de terras férteis com a irrigação e os minerais que adubam a terra; negativamente pela negação da frase anterior ou pelo seu excesso, que impedem uma atividade agrícola normal. Foi feita uma análise Kmeans a esta variável no *paper* de Parracho, Melo Gonçalves, & Rocha, (2015). Os resultados obtidos foram 6 centroides que foram utilizados para calcular tendências nos índices de precipitação usando a regressão de Theil-Sen. Isto resultou em seis regiões com funções densidade de probabilidade estatisticamente diferentes para a precipitação nos centróides dos *clusters*. A unidade de medida utilizada é o milímetro – mm;
- Temperatura média – frio ou calor em excesso são inóspitos e também impedem uma atividade agrícola produtiva e variada. O sector primário tinha um grande peso no período pré-industrial e ajudava muito a definir as localizações. A unidade de medida é grau Celsius multiplicado por 10 – 10* °C.

5.1.1. Evolução da análise e resultados obtidos

Os resultados que abaixo se transcrevem são fruto de uma análise pormenorizada para os dois períodos temporais em questão. Obedeceram a uma escolha de variáveis e de regiões que abaixo se descreve.

Optou-se sempre por fazer uma análise de 2 a 10 *clusters*. Dez é um valor elevado que obrigaria, caso fosse este o resultado, a uma análise mais aprofundada. Apesar de ser possível, posteriormente, fazer regressões dentro de cada *cluster* (pois o valor inicial de observações é de 12130), existiriam “nichos de mercado” ou, no caso vertente, nichos de polígonos.

Convém recapitular que, dos 12130 polígonos iniciais, estamos agora a trabalhar com 11989 após a exclusão das 142 unidades administrativas sem população.

5.1.1.1. 1878-1940

Ao incluir todas as variáveis acima listadas (com exceção da curvatura média), o melhor resultado são 7 *clusters*. Tendo em conta que a correlação entre o declive e a sua variabilidade tem um valor de 0,9724, decidimos excluir a variável variabilidade do declive. Fazendo uma análise de *clusters* com as variáveis TCAM18781940, Altitude, Curvatura, Temperatura e Precipitação Médias, obtemos quatro *clusters*.

Colocou-se a hipótese de que a província de Madrid poderia constituir um *cluster* à parte, tratado como *outlier*, pelos seguintes motivos: por um lado, a província de Madrid tem municípios que tiveram um grande crescimento populacional comparando com a totalidade da Península Ibérica

(Tabela 4); por outro lado, esta metrópole não se localiza junto ao litoral, ao contrário das outras grandes áreas urbanas. Acresce ainda que os polígonos da província de Madrid, devido à sua localização central na Península Ibérica, têm uma altitude média de 569 metros (Tabela 5). O cruzamento de elevadas taxas de crescimento, com altitude média relativamente elevada, poderia resultar na existência de uma província *outlier*.

Tabela 4 - Taxas de Crescimento da Província de Madrid e restantes Regiões

Taxa de Crescimento do Período	Madrid	Restantes Regiões	Península Ibérica
1878-1940	164%	55%	58%
1940-2001	245%	39%	49%
1878-2001	809%	115%	135%

Tabela 5 - Média de Altitude por Região

Região Administrativa	Média de Altitude por Região (em metros)
Madrid	569,175
Restante Península	849,183
Total Geral	573,330

Efetivamente, em países com faixa costeira, as taxas de crescimento populacional elevadas ocorrem em áreas metropolitanas junto ao litoral (Rappaport & Sachs, 2003). Na Península Ibérica isto acontece em Lisboa, Porto, Valencia, Múrcia, Bilbao, Sevilha e Barcelona. Comparando os mapas de *clusters* obtidos com e sem a província de Madrid para os dois períodos (Figura 4 - Mapa dos *clusters* obtidos sem os dados referentes à província de Madrid no período 1878-1940 e Figura 5 respetivamente), chegou-se à conclusão de que a província se insere num contínuo de três *clusters*. Na província de Toledo, a sul de Madrid, há também municípios em dois *clusters* com ligação a Madrid. Nas províncias de Segóvia, a Norte, de Guadalajara, a Leste, e de Ávila, a Oeste, há municípios em maioritariamente dois *clusters* na mesma situação. Assim sendo, optou-se pelos quatro *clusters* com a província de Madrid para esta janela temporal (Figura 5).

Figura 4 - Mapa dos *clusters* obtidos sem os dados referentes à província de Madrid no período 1878-1940

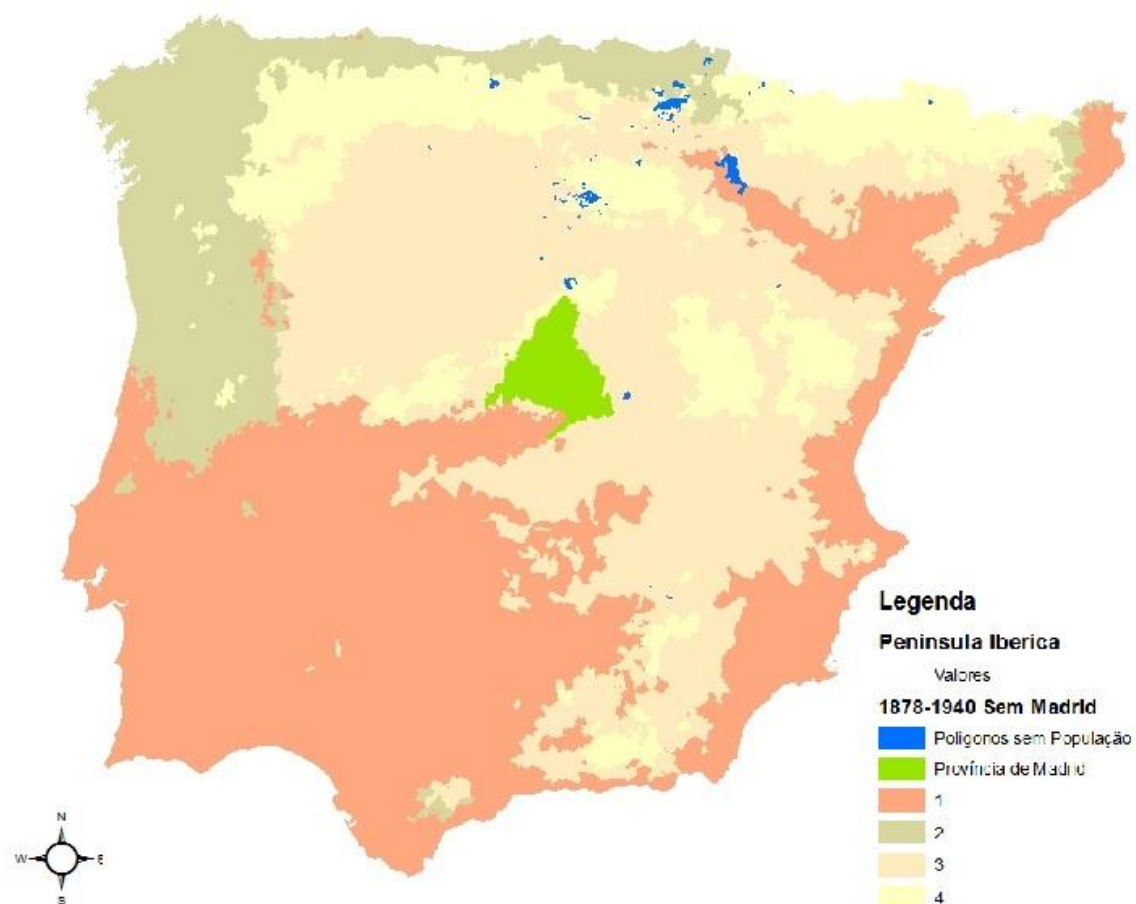
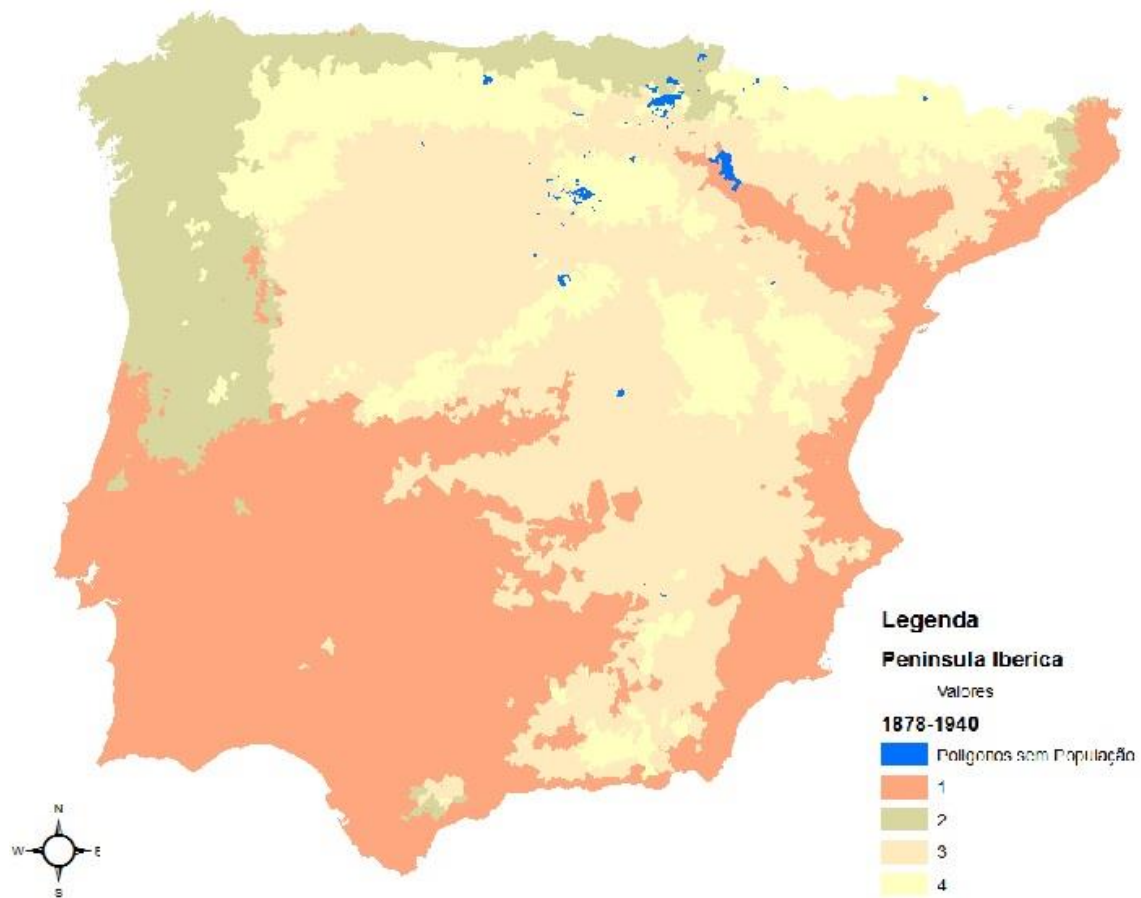


Figura 5 - Mapa dos *clusters* obtidos com os dados referentes à província de Madrid no período 1878-1940



5.1.1.2. 1940-2001

Na análise a este período foram feitas várias experiências que obedeceram aos raciocínios e regras utilizados para as décadas anteriores à Segunda Guerra Mundial. Utilizando todas as variáveis em causa, à exceção da curvatura média, obtivemos um resultado com 10 *clusters*. Retirando a variabilidade do declive, o número de *clusters* que maximiza o critério CCC passa a ser dois. No entanto, sendo este um valor demasiado baixo, passamos ao número ótimo de *clusters* seguinte, que é quatro.

Novamente se colocou a questão da província de Madrid e, novamente pelas mesmas razões, se decidiu incluir os municípios desta província na análise (Figura 6 e Figura 7).

Figura 6 - Mapa dos *clusters* obtidos sem os dados referentes à província de Madrid no período 1940-2001

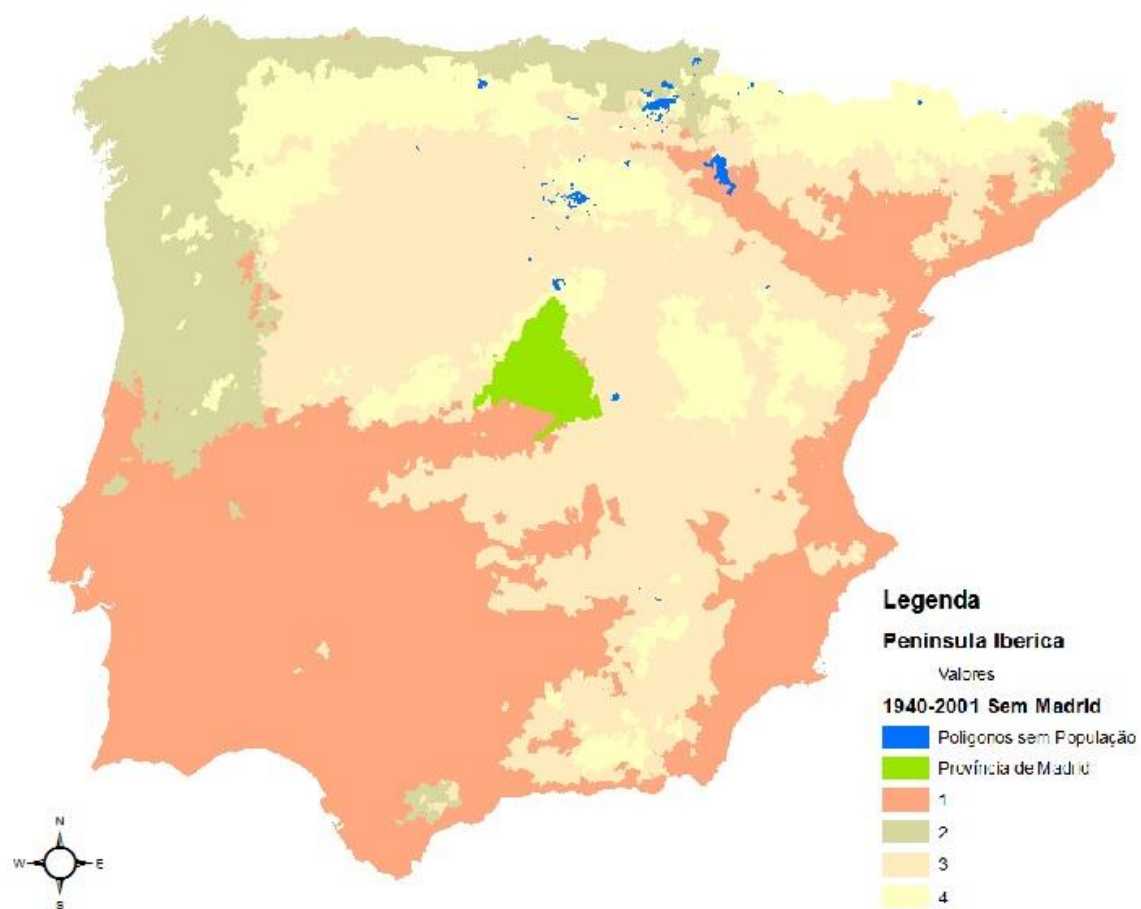
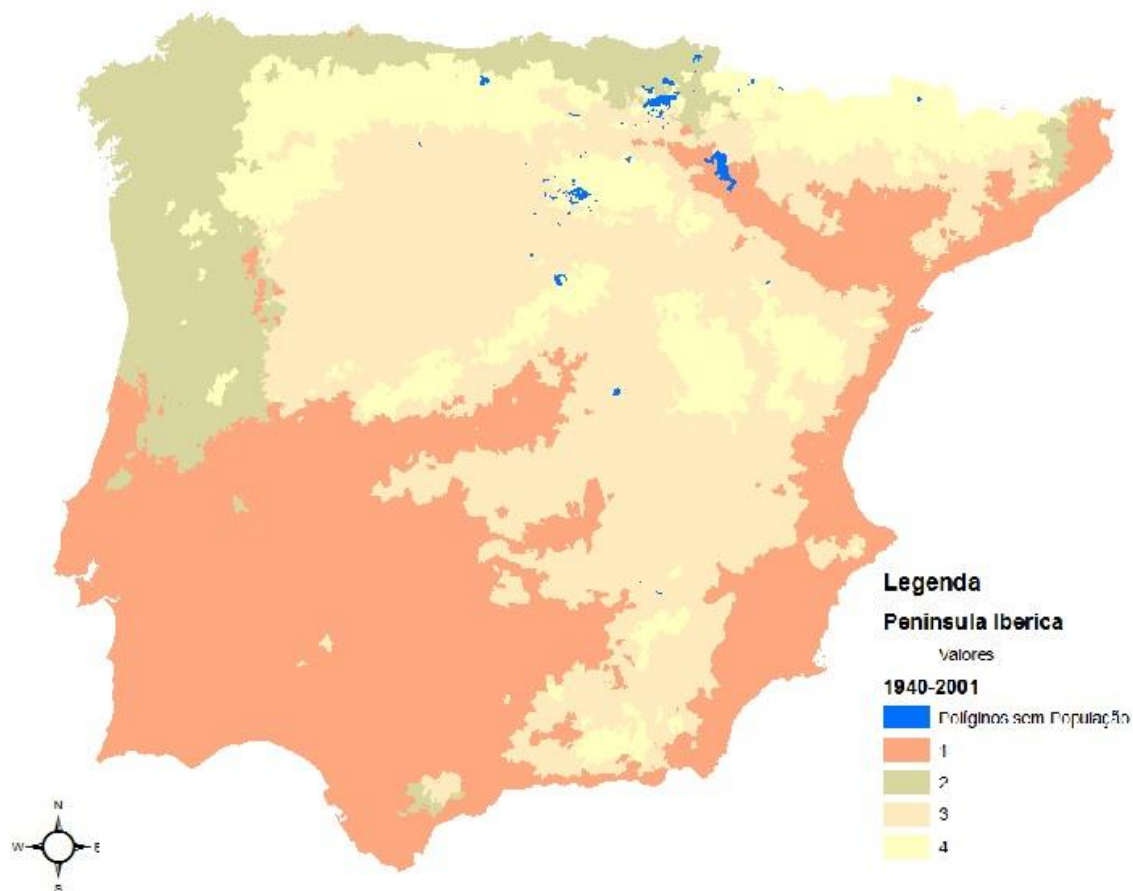


Figura 7 - Mapa dos *clusters* obtidos com os dados referentes á província de Madrid no período 1940-2001



5.1.2. Caracterização dos Segmentos

Observando com mais atenção a Figura 5 e a Figura 7 notam-se algumas mudanças de *cluster*, principalmente nas regiões de Madrid, Toledo, Navarra, Ciudad Real, Múrcia, Valencia Barcelona, Lleida e Tarragona. Excluindo estas mudanças, que se incluem no grupo que será abaixo escalpelizado e que totaliza 3,3% dos polígonos em análise, com uma população de 883.714 habitantes em 1940 e 2.932.496 habitantes em 2001, há uma manutenção de polígonos nos *clusters* definidos para os dois períodos. Passemos primeiramente à análise dos resultados numéricos obtidos para os dois períodos.

Para facilitar a análise, e dada a semelhança dos segmentos nos dois períodos, a nomenclatura dos *clusters* tem no algarismo das dezenas o número dos clusters no período 1878-1940 e no algarismo das unidades o número dos clusters no período 1940-2001. Os resultados da segmentação (Figura 5, Figura 7, Figura 8 e Figura 9 e Tabela 6 e Tabela 7) são resumidos em seguida.

Tabela 6 - Quadro com Valores Médios de cada Variável dentro de cada *cluster* e número de polígonos desse *cluster* relativos ao período 1878-1940

Cluster	TCAM18781940	ALT_MEAN	SLP_MEAN	WCPRECMEAN	WCTEMPMEAN	N.º de Polígonos
1	0,8044	246,4487	3,0927	595,5924	159,1199	3 319
2	0,4674	378,9453	6,6256	1 158,8437	135,0084	3 365
3	0,2439	796,4865	3,3743	520,7178	120,7149	4 028
4	-0,0444	1 194,0525	10,0505	760,5565	93,7216	1 277

Figura 8 - Gráfico de Coordenadas Paralelas de cada Variável dentro de cada *cluster* e número de polígonos desse *cluster* relativos ao período 1878-1940

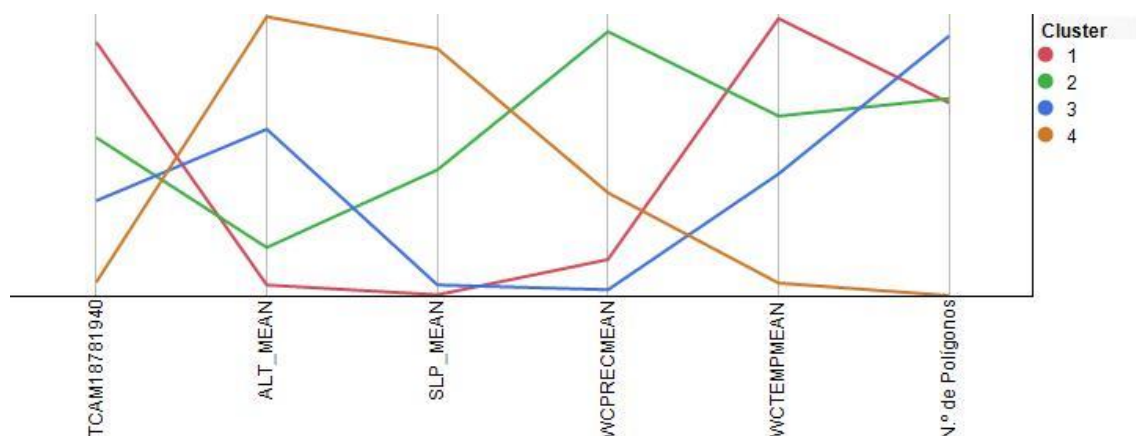
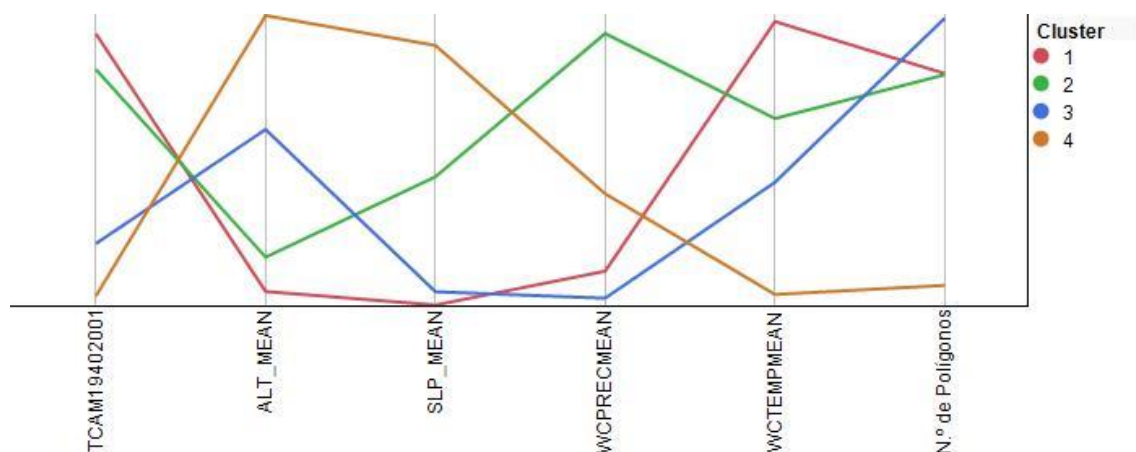


Tabela 7 - Quadro com Valores Médios de cada Variável dentro de cada *cluster* e número de polígonos desse *cluster* relativos ao período 1940-2001

Cluster	TCAM19402001	ALT_MEAN	SLP_MEAN	WCPRECMEAN	WCTEMPMEAN	N.º de Polígonos
1	0,1077	256,4631	3,0812	587,9928	158,4386	3 401
2	-0,1668	373,4649	6,5822	1 155,7665	135,2781	3 388
3	-1,5057	809,5520	3,4464	523,4231	120,0342	3 968
4	-1,9096	1 198,3233	10,1712	773,1484	93,2443	1 232

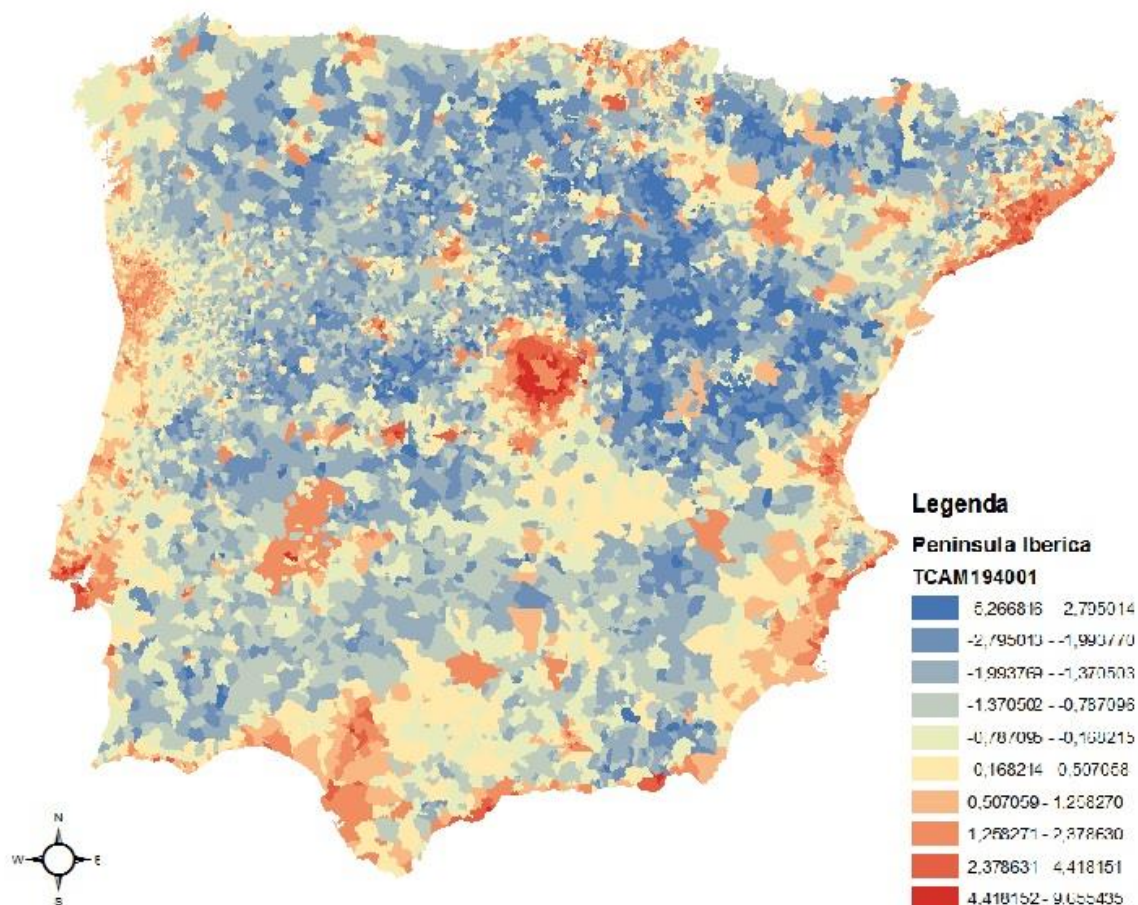
Figura 9 - Gráfico de Coordenadas Paralelas de cada Variável dentro de cada *cluster* e número de polígonos desse *cluster* relativos ao período 1940-2001



Há uma clara divisão entre o Norte e o Sul de Portugal. O Norte está no *cluster* 22, que faz um contínuo com a Galiza, Astúrias, Cantábria e País Basco. A divisão Norte / Sul é uma divisão que se faz no vale do Rio Tejo, desde a fronteira até aos concelhos do litoral do Distrito de Leiria. Estes, ao contrário dos restantes, incluem-se no Sul de Portugal. A vertente sudoeste do Parque Natural da Serra das Neves, parte do Parque Natural das Serras de Aire e dos Candeeiros e uma zona entre Vic e Girona, são as ilhas deste território. Observando a Taxa de Crescimento Anual Médio, este *cluster* fica nos dois períodos na 2.ª posição, decrescendo de 46,71% para -16,68%. Os valores geográficos médios sofrem ligeiras alterações de um período para o outro. É uma região com polígonos de pequena dimensão em que a maior área metropolitana é a do Porto.

O Sul de Portugal, acima definido, faz um contínuo com a Extremadura e parte ocidental da Andaluzia, seguindo a autoestrada n.º 5 espanhola até Madrid – *cluster* 11. Engloba ainda a esmagadora maioria da costa sul e leste da Península, sendo que a penetração para o interior é superior na costa leste. Estende-se ainda pelo vale do Rio Ebro – zona com uma Taxa de Crescimento Anual Médio elevada no 2.º período (vide Figura 10 - Mapa de decis da Taxa de Crescimento Anual Médio 1940-2001). À exceção da Área Metropolitana do Porto, é neste *cluster* que estão todas as grandes cidades da Península – Lisboa, Madrid, Barcelona e Valência. Esta região tem as maiores Taxas de Crescimento Anual Médio nos dois períodos e é a única com crescimento positivo em ambos os períodos.

Figura 10 - Mapa de decis da Taxa de Crescimento Anual Médio 1940-2001



O *cluster* 33 situa-se quase na sua totalidade no interior da Península Ibérica e em Espanha, embora exista uma visível faixa fronteiriça de freguesias em Portugal, nos concelhos raianos de Almeida, Figueira de Castelo Rodrigo e Sabugal, no distrito da Guarda, e em todos os concelhos raianos de Bragança e ainda em Macedo de Cavaleiros, Alfandega da Fé e Torre de Moncorvo, no mesmo distrito. Em Espanha, é maioritariamente constituído por municípios de Castela la Mancha e Castela Leon e províncias de Jaén, Granada e Almería na Andaluzia. A TCAM desce de 24,39% para -150,27% com uma altitude média a rondar os 800 metros, declive 3.40, precipitação média de 520 mm e temperatura média de 12 °C.

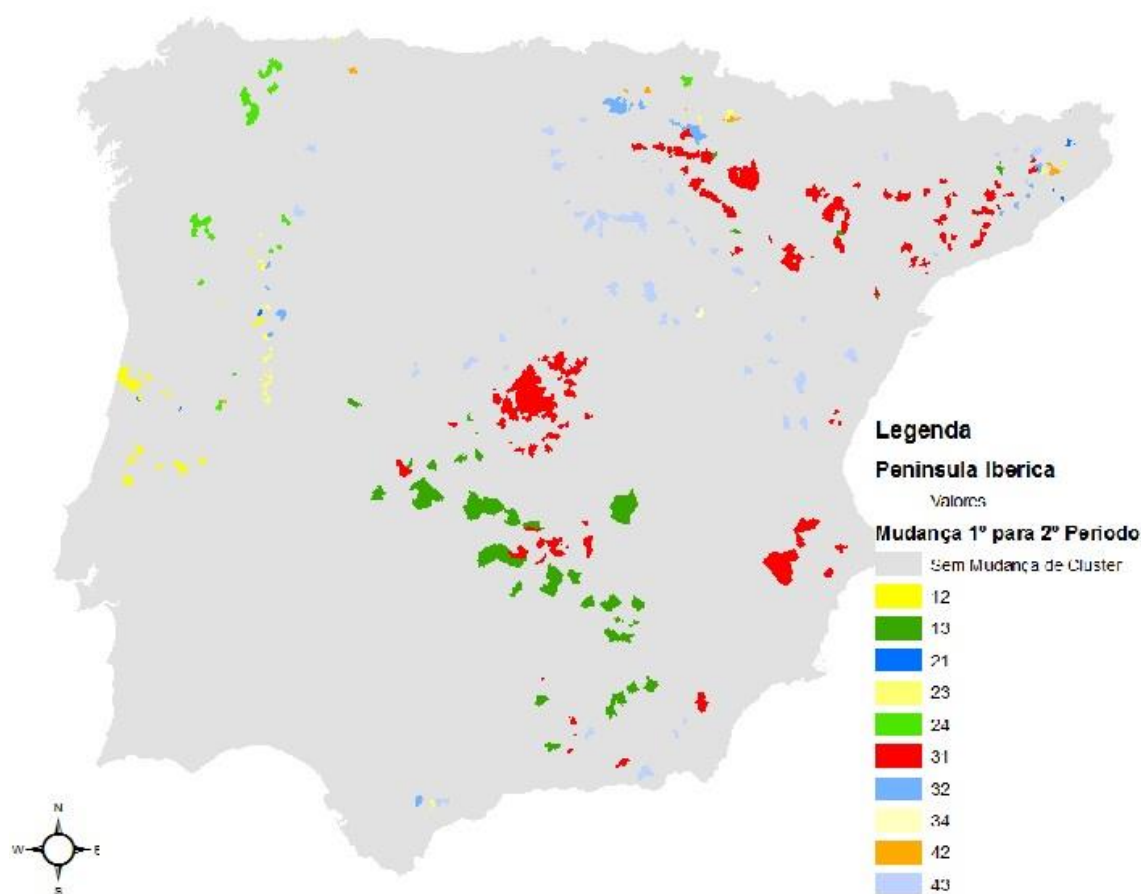
O *cluster* 44 engloba as Serras e Parques Naturais – em Espanha, todos os Picos da Europa, Serra de Guadarrama (a norte de Madrid), Parque de Montesinho (em Bragança), que continua em Espanha, Parque Natural do Alto Tejo, Parque Natural da Serra da Cebollera, Parque Natural Serras de Cazorla, Segura y Las Villas, Parque Natural Serra de Baza e Serra Nevada. São zonas inóspitas, com baixas temperaturas médias e alta pluviosidade, rugosas e com altitude elevada. Os parques naturais foram decretados enquanto tal nalgumas destas regiões devido à baixa taxa de “humanização” e às condições naturais existentes, consideradas excecionais. Se o seu crescimento populacional era negativo no 1.º período e o valor o mais baixo da TCAM desse período, esse decréscimo acentuou-se no segundo período mantendo o *ranking* no segundo período.

A fronteira política entre Portugal e Espanha desaparece mais uma vez, tal como no trabalho de Da Silveira et al., (2013). Há zonas contínuas, tal como descrito em epígrafe. Apesar disso, há diferenças entre os dois países. Portugal tem mais de 90% dos polígonos concentrados em dois *clusters*, enquanto Espanha precisa de 3 *clusters* para totalizar 85% dos polígonos.

5.1.3. Comparação dos dois períodos

A observação das mudanças de *cluster* leva-nos a concluir, a partir da Figura 11 - Mapa das mudanças de *cluster* entre 1878-1940 e 1940-2001, o seguinte:

Figura 11 - Mapa das mudanças de *cluster* entre 1878-1940 e 1940-2001



- Mudança 13 – polígonos que têm a TCAM mais alta em 1878-1940 passam para a 3.ª posição em 1940-2001 – a média passa de 139% para -142%. São maioritariamente as províncias de Ciudad Real, Granada, Jaén e Toledo. Se no 1.º período tem a taxa de crescimento mais alta, no segundo é nitidamente negativa. Passa de 71.997 habitantes em 1878 para 180.734 em 1940 e este valor é drasticamente reduzido para 116.294 em 2001 (vide Tabela 8);
- Mudança 31 – polígonos que refletem a urbanização dos grandes centros urbanos espanhóis de Madrid (com Toledo), Barcelona (com Tarragona e Lleida), Valencia (com Alicante), Navarra e Saragoça. A população sextuplicou de 1940 para 2001 (vide Tabela 8), o que é corroborado pela TCAM mais elevada nos grupos definidos;
- Mudança 43 – também neste grupo pode existir alguma perda de influência dos fatores geográficos e do clima na distribuição da população com o 2.º maior número de mudança de polígonos: 59. A população mantém-se nestes polígonos, o que não acontece no mais inóspito (44) em que quase desce para metade.

As três mudanças anteriores ocorrem exclusivamente em território espanhol.

- Mudança 12 – polígonos que descem da 1.ª posição para a 2.ª posição no segundo período. Vinte e cinco destes 41 polígonos situam-se no concelho de Vagos, no distrito de Aveiro, e nos concelhos circundantes. Há uma perda de influência desta parte sul do distrito de Aveiro. Sendo a

cidade de Aveiro o limite sul da Área Metropolitana do Porto, é natural que as populações se deslocassem mais para norte. Os restantes casos, nos quais se incluem os dois únicos em território espanhol, são casos isolados.

- Mudança 21 – Polígonos que sobem da 2.ª posição para a 1.ª posição no segundo período. São apenas seis polígonos, três de cada lado da fronteira. Refletem casos isolados em Portugal e pouco prováveis de acontecer. Em Espanha, são na região de Barcelona e são a passagem da ilha do *cluster* 2, em 1878-1940, para o *cluster* 1 em 1940-2001 – espelham a urbanização dos grandes centros urbanos.
- Mudança 23 – Este *downgrade*, que ocorre 83% em território português, mais precisamente nos distritos de Bragança e Guarda, é a demonstração da desertificação do interior luso. Freguesias dos concelhos raianos de Almeida e Sabugal predominam neste grupo.
- Mudança 24 – Este *downgrade* acentuado ocorre nos interiores transmontano, galego e asturiano. São casos mais agravados da mudança 23. Não há uma localização predominante.
- Mudança 32 – Nas ocorrências de Navarra e Álava, são uma subida de nível até ao segmento dominante local – o segundo ou 22. Nas ocorrências de Barcelona, é o sair de um certo interior para uma urbanização menos intensa.
- Mudança 42 – casos pontuais acentuados da mudança 32.
- Mudança 34 – restringe-se a dois polígonos e é irrelevante em termos populacionais.

Tabela 8 - Quadro com Somatório de População por *Cluster* de origem e de chegada para os censos de 1878, 1940 e 2001

Cluster de Origem / Destino	Soma de POPD1878	Soma de POPR1940	Soma de POPR2001 (1)	Previsão 2001 (2)	Erro da Previsão 2001 (3) = (2) - (1)
11	8 241 480	14 863 505	25 585 817	22 084 501	3 501 316
12	41 900	82 458	92 540	122 518	-29 978
13	71 997	180 734	116 294	268 538	-152 244
21	3 803	3 597	6 653	5 344	1 309
22	5 367 201	8 071 199	10 879 508	11 992 353	-1 112 845
23	13 571	16 214	4 572	24 091	-19 519
24	33 158	41 696	14 074	61 953	-47 879
31	309 587	378 731	2 306 140	562 726	1 743 414
32	69 142	93 393	300 225	138 765	161 460
33	4 735 444	6 962 382	7 791 407	10 344 850	-2 553 443
34	820	957	90	1 422	-1 332
42	11 627	11 638	15 722	17 291	-1 569
43	75 133	74 296	76 186	110 391	-34 205
44	1 439 162	1 553 288	853 424	2 307 907	-1 454 483
Total Geral	20 414 025	32 334 089	48 042 652	48 042 652	0

Neste quadro, é de referir a metodologia a que obedeceu a construção da coluna Previsão 2001. Aplicou-se a Taxa de Crescimento do Total Geral da Península Ibérica $(48.042.652/32.334.089-1)*100$ de 1940 a 2001 aos valores da população em 1940. Esta coluna teve como objetivo, mantendo os pesos relativos de todos os clusters, ter um ponto de referência ao qual se podem comparar os dados reais para o 2º período

Analisando a Matriz de transferências de polígonos (Tabela 9), temos 259 *upgrades* (upgrade corresponde à mudança de cluster subindo na escala de TCAM entre os dois períodos) e 132 *downgrades*. Estes valores são reveladores de uma aglomeração da população, como se pode constatar na Tabela 10, semelhante à Matriz de transferências de polígonos mas com as densidades populacionais em 2001.

Tabela 9 - Matriz de transferências de polígonos em termos de alterações na TCAM entre os dois períodos

Número de Polígonos		Clusters 1940-2001				Total
		1	2	3	4	
Clusters 1878-1940	1	3 233	41	45		3 319
	2	6	3 315	24	20	3 365
	3	162	24	3 840	2	4 028
	4		8	59	1 210	1 277
Total		3 401	3 388	3 968	1 232	11 989

Células a verde - *upgrade*

Células a vermelho - *downgrade*

Células a amarelo - manutenção

Tabela 10 - Densidade Populacional da Matriz de transferências de polígonos da Tabela anterior

Densidade Populacional em 2001		Clusters 1940-2001				Total
		1	2	3	4	
Clusters 1878-1940	1	599	142	44		586
	2	64	343	11	13	338
	3	371	360	29	2	45
	4		107	22	10	11
Total		587	340	29	10	270

Células a verde - *upgrade*

Células a vermelho - *downgrade*

Células a amarelo - manutenção

Apesar de os segmentos definidos terem demonstrado a continuidade entre os dois lados da fronteira, as mudanças de *cluster* não ocorreram de forma simétrica, o que se pode comprovar pelo quadro seguinte recorrendo às Regras de Associação:

Total de Polígonos	11.989
Espanha	7.952
Mudança de Cluster	391
Espanha + Mudança de Cluster	311

Confiança	3,91%
Suporte	2,59%
Confiança Esperada	3,26%
<i>Lift</i>	1,20

O valor do *Lift* prova que um polígono espanhol é 1,20 vezes mais provável de mudar de *cluster* do que um polígono português.

5.2. REGRESSÃO ESPACIAL

Como seria de esperar, para todas as regressões exploratórias verificou-se que há multicolinearidade entre os vários registos. O número de condição de multicolinearidade demonstra isso ao termos sempre valores superiores a 30. Todos os testes de I de moran demonstram isso.

Para ser feita inferência estatística a regressão tem de passar no teste de Jarque-Bera da normalidade dos resíduos. No entanto isto não aconteceu em qualquer das regressões dos mínimos quadrados ordinários feita.

Já o teste de Breusch-Pagan à heteroscedasticidade dos resíduos, aplicável aos modelos spatial lag model e spatial error model apresentou resultados homoescedásticos em dois modelos.

Como abordagem exploratória fizemos os testes de diagnóstico de dependência espacial para tentar perceber qual o modelo mais adequado a cada cluster.

Para fazer as regressões ao utilizar o programa GeoDa foi aconselhado por motivos de qualidade dos resultados evitar o uso de polígonos sem qualquer vizinho. Ao tomar esta opção excluíram-se os polígonos constantes na Tabela 41. Comparando o critério AIC das regressões dos mínimos quadrados ordinários, com e sem estes polígonos, para os dois períodos em análise e para os quatro clusters gerados, os resultados foram sempre favoráveis às regressões sem estes polígonos, como se pode observar na Tabela 11 (o significado da abreviatura atribuída a cada regressão realizada encontra-se na Tabela 3 do capítulo 4).

Tabela 11 - Tabela comparativa do critério AIC com e sem polígonos isolados

Regressão	Com Isolados (1)	Sem Isolados (2)	Diferença (3) = (1) - (2)
T1A1S1L0R0T1P1CL1	7 828.86	7 764.96	63.90
T1A1S1L0R0T1P1CL2	5 015.31	4 985.19	30.12
T1A1S1L0R0T1P1CL3	5 604.96	5 499.40	105.56
T1A1S1L0R0T1P1CL4	1 796.87	1 754.65	42.22
T1A1S1L1R1T1P1CL1	7 804.93	7 741.31	63.62
T1A1S1L1R1T1P1CL2	5 015.67	4 985.47	30.20
T1A1S1L1R1T1P1CL3	5 488.02	5 378.20	109.82
T1A1S1L1R1T1P1CL4	1 774.39	1 734.71	39.68
T2A1S1L0R0T1P1CL1	11 584.50	11 502.10	82.40

T2A1S1L0R0T1P1CL2	8 929.07	8 868.70	60.37
T2A1S1L0R0T1P1CL3	13 059.50	13 000.80	58.70
T2A1S1L0R0T1P1CL4	3 731.78	3 664.90	66.88
T2A1S1L1R1T1P1CL1	11 383.20	11 302.30	80.90
T2A1S1L1R1T1P1CL2	8 813.18	8 763.33	49.85
T2A1S1L1R1T1P1CL3	12 942.20	12 883.00	59.20
T2A1S1L1R1T1P1CL4	3 705.79	3 639.00	66.79

5.2.1. Período 1878-1940

5.2.1.1. Cluster 1

Neste período e neste *cluster*, os 4 testes de dependência espacial não são esclarecedores para a escolha entre o *Spatial Lag Model* ou para o *Spatial Error Model*, como se pode ver na Tabela 12. Acrescentar as variáveis Distância ao Litoral e Distância aos Rios Principais não traz nenhuma melhoria nesta análise, como se pode verificar na Tabela 13. Como todos os testes de dependência espacial são significativos, passámos à realização do *Spatial Error Model* e do *Spatial Lag Model*.

Tabela 12 - Diagnóstico de Dependência Espacial da Regressão T1A1S1L0R0P1T1CL1OLS

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.3866	60.2578	0
Lagrange Multiplier (lag)	1	3384.4655	0
Robust LM (lag)	1	65.7072	0
Lagrange Multiplier (error)	1	3540.3559	0
Robust LM (error)	1	221.5975	0
Lagrange Multiplier (SARMA)	2	3606.0630	0

Tabela 13 - Diagnóstico de Dependência Espacial da Regressão T1A1S1L1R1P1T1CL1OLS

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.3869	60.5734	0
Lagrange Multiplier (lag)	1	3378.7238	0
Robust LM (lag)	1	40.7070	0
Lagrange Multiplier (error)	1	3546.7391	0
Robust LM (error)	1	208.7223	0
Lagrange Multiplier (SARMA)	2	3587.4461	0

Na Tabela 14 podemos ver os coeficientes e resultados dos testes do *Spatial Error Model* feito. A variável Altitude Média não é significativa.

Tabela 14 - Coeficientes e testes da Regressão T1A1S1L0R0P1T1CL1SEM

Variable	Coefficient	Std.Error	z-value	Probability
CONSTANT	-2.755530000	1.0645380000	-2.588474	0.00964
ALT_MEAN	0.000210884	0.0002973877	0.7091227	0.47825
SLP_MEAN	-0.036056490	0.0077141340	-4.674082	0.00000
WCPRECMEAN	0.000786148	0.0003323142	2.365677	0.01800
WCTEMPMEAN	0.019786330	0.0057114280	3.464341	0.00053
LAMBDA	0.810565500	0.0155638900	52.07989	0.00000

Este Modelo revela heteroscedasticidade.

Refazendo a regressão sem a variável Altitude Média chegamos aos valores vertidos para a Tabela 15.

Tabela 15 - Coeficientes e testes da Regressão T1A0S1L0R0P1T1CL1SEM

Variable	Coefficient	Std.Error	z-value	Probability
CONSTANT	-2.134553000	0.5991537000	-3.562613	0.00037
SLP_MEAN	-0.034301940	0.0073835440	-4.645729	0.00000
WCPRECMEAN	0.000723016	0.0003232889	2.23644	0.02532
WCTEMPMEAN	0.016413000	0.0031123420	5.273521	0.00000
LAMBDA	0.812681200	0.0154346300	52.65312	0.00000

A heteroscedasticidade continua presente.

Na Tabela 16 podemos ver os resultados da regressão *Spatial Lag Model* executada.

Tabela 16 - Coeficientes e testes da Regressão T1A1S1L0R0P1T1CL1SLM

Variable	Coefficient	Std.Error	z-value	Probability
W_TCAM187840	0.786864100	0.0162203600	48.51089	0.00000
CONSTANT	-0.955359000	0.2399185000	-3.982014	0.00007
ALT_MEAN	0.000169783	0.0000808017	2.101236	0.03562
SLP_MEAN	-0.036726510	0.0046356600	-7.922608	0.00000
WCPRECMEAN	0.000212635	0.0000786525	2.703478	0.00686
WCTEMPMEAN	0.006718155	0.0013260150	5.066426	0.00000

A heteroscedasticidade parece ser omipresente qualquer que seja a regressão feita.

Ao prosseguirmos com a análise acrescentando as variáveis distância ao litoral e distância aos rios principais continuamos a ter heteroscedasticidade.

Uma das soluções para os erros apresentados poderia ser aumentar o número de variáveis, o que não é possível no caso presente. Procedeu-se, então, à execução de uma *Geographic Weighted Regression*.

Na GWR há um erro de especificação do modelo com as quatro variáveis independentes. Fazendo GWR para cada uma das variáveis, só não se consegue obter resultados com a Temperatura Média. Quando este conjunto de resultados acontece deve-se excluir a variável não explicativa de *per si*. Fazendo uma regressão com as três restantes variáveis obtém-se uma regressão válida. Verificou-se se a logaritmação da temperatura média poderia trazer esta variável de volta à análise, mas o modelo gerado tem o mesmo tipo de erro de especificação que tinha com a variável original.

Os parâmetros dos coeficientes para as três variáveis explicativas seguem uma distribuição quase contínua, como se pode ver nas Figura 12, Figura 13 e Figura 14. Os valores são positivos ou negativos, conforme o polígono, e não temos um valor único para todo o segmento tal como acontecia nos modelos anteriores.

Figura 12 - Coeficientes da variável Altitude Média por polígono da regressão GWRT1A1S1R0L0P1T0CL1

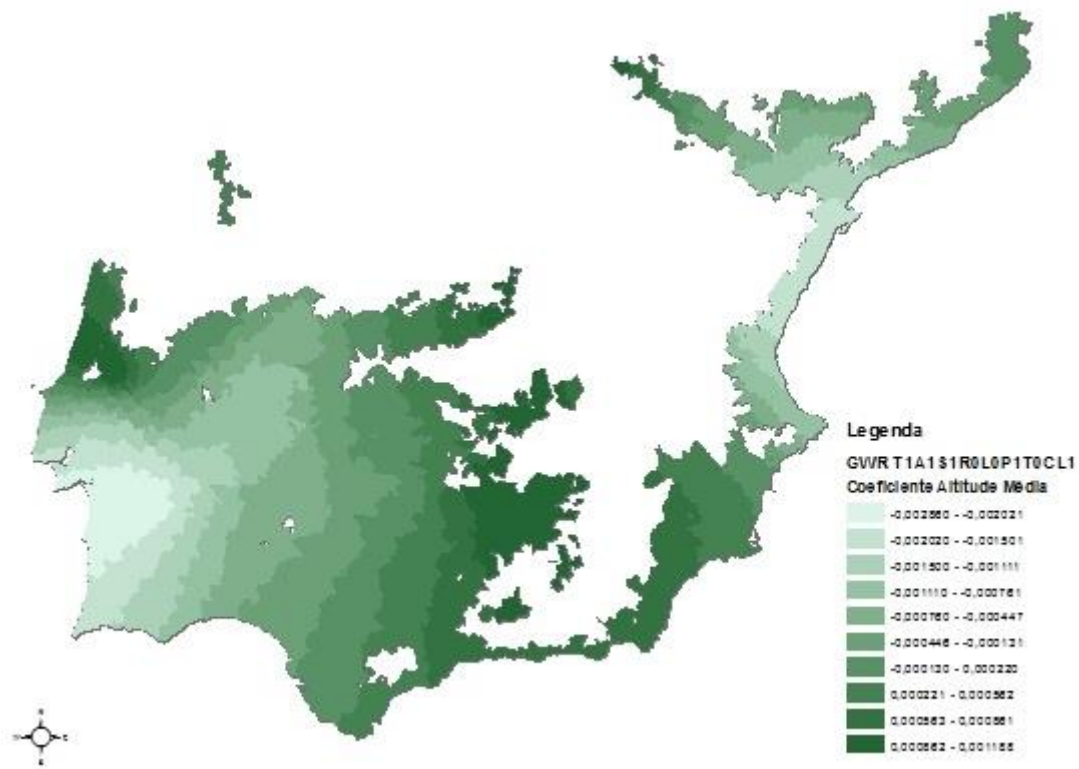


Figura 13 - Coeficientes da variável Declive Médio por polígono da regressão GWRT1A1S1R0L0P1T0CL1

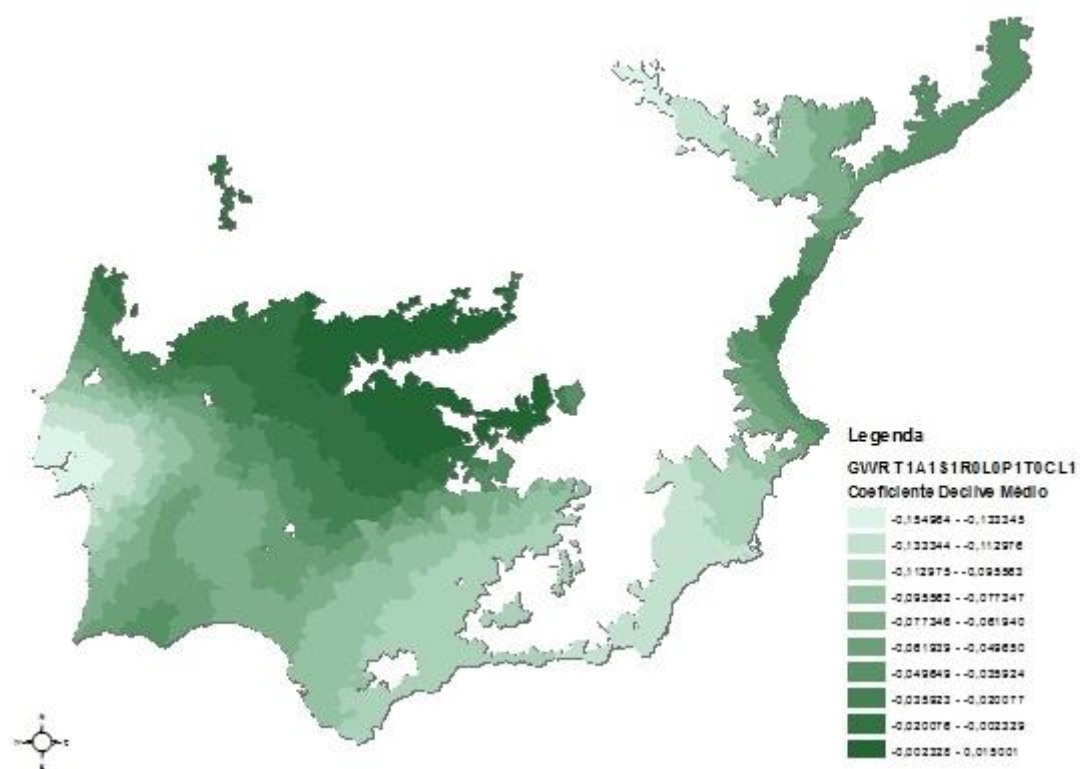
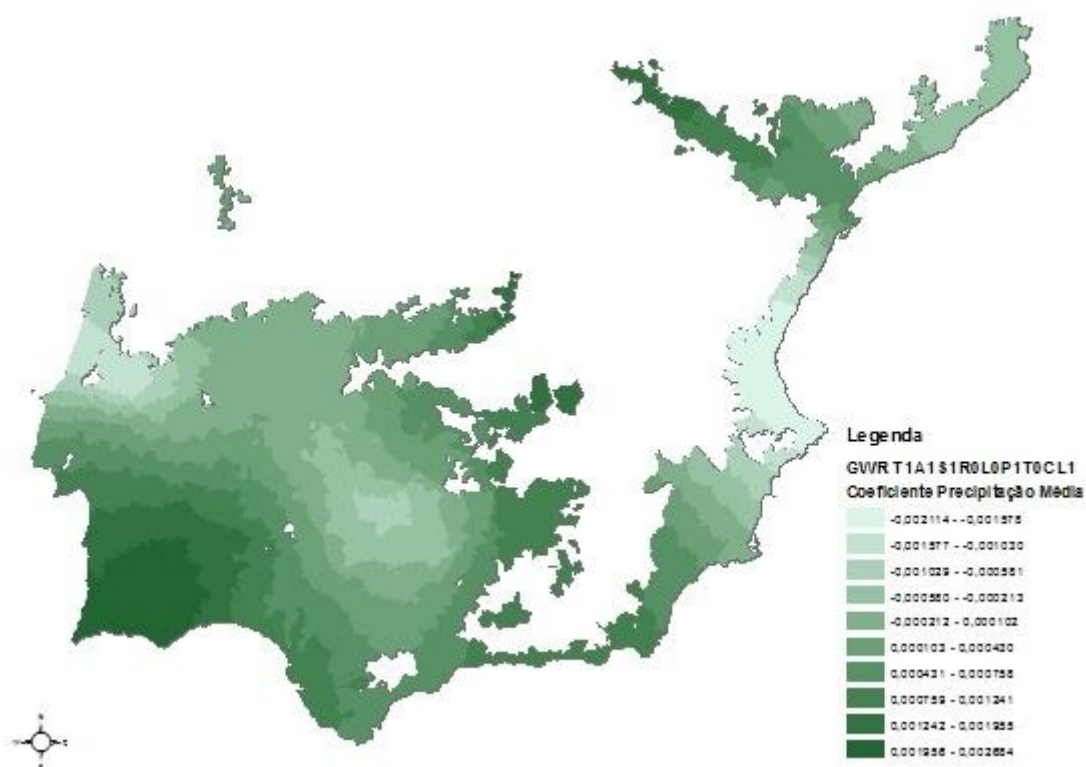
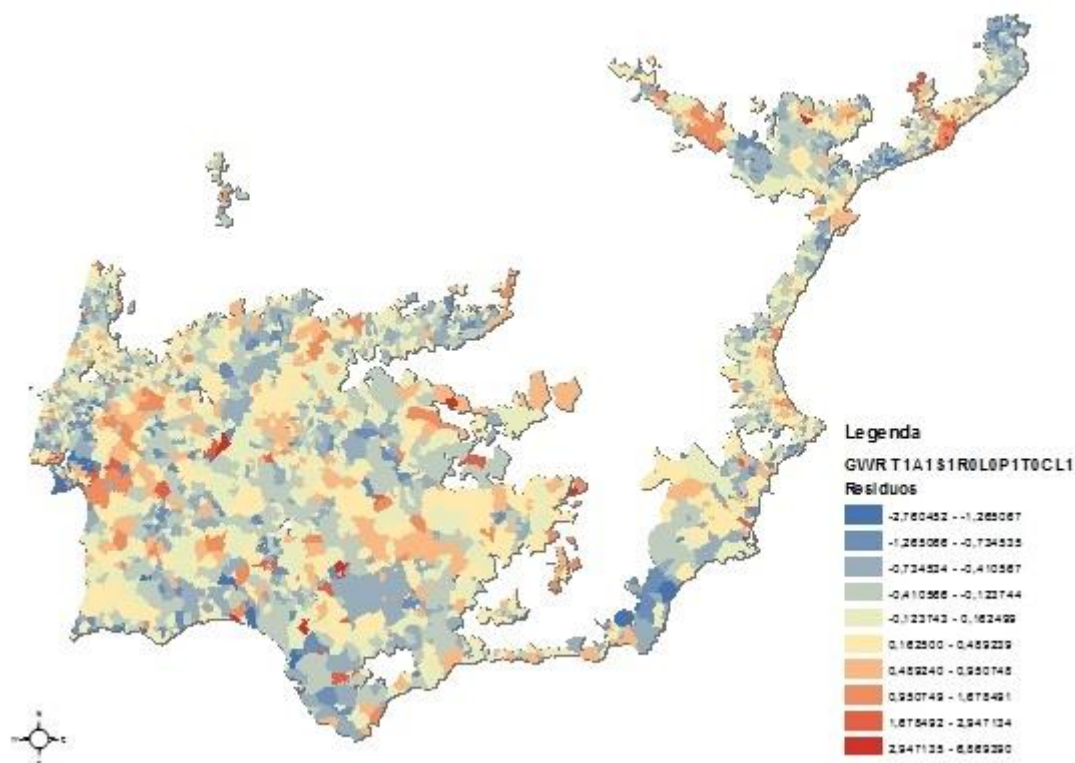


Figura 14 - Coeficientes da variável Precipitação Média por polígono da regressão GWRT1A1S1R0L0P1T0CL1



Os Resíduos não são homogêneos como é suposto ser quando não há autocorrelação espacial, como é patente na Figura 15.

Figura 15 - Resíduos da regressão GWRT1A1S1R0L0P1T0CL1



Os coeficientes apresentam valores negativos e positivos consoante o polígono em causa. Há uma percentagem diferente de valores positivos e negativos para o terno de variáveis. A precipitação influencia positivamente cerca de metade da TCAM dos polígonos, enquanto a altitude só os influencia em 30% e o declive médio apenas 10% dos polígonos. Este valor está em sintonia com a hipótese de que o declive elevado torna um terreno inóspito.

O critério AIC é mais baixo no primeiro período (1878-1940) do que no segundo (1940-2001), como seria de esperar, de acordo com a hipótese da perda de influência das variáveis geográficas na distribuição da população na Península Ibérica. O AIC sobe em relação aos melhores modelos obtidos em SEM e SLM.

5.2.1.2. Cluster 2

O mesmo sucedeu na 1.ª fase de despistagem deste *cluster*. À semelhança do *cluster* 1, a dependência espacial é estatisticamente significativa, passámos à realização do *Spatial Error Model* e do *Spatial Lag Model*, como se pode ver nas Tabela 17 e Tabela 18.

Tabela 17 - Diagnóstico de Dependência Espacial da Regressão T1A1S1L0R0P1T1CL2OLS

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.3791	62.3215	0.00000
Lagrange Multiplier (lag)	1	3590.2943	0.00000
Robust LM (lag)	1	20.4041	0.00001
Lagrange Multiplier (error)	1	3784.3049	0.00000
Robust LM (error)	1	214.4146	0.00000
Lagrange Multiplier (SARMA)	2	3804.7090	0.00000

Tabela 18 - Diagnóstico de Dependência Espacial da Regressão T1A1S1L1R1P1T1CL2OLS

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.3786	62.5718	0.00000
Lagrange Multiplier (lag)	1	3585.8497	0.00000
Robust LM (lag)	1	20.5343	0.00001
Lagrange Multiplier (error)	1	3775.7407	0.00000
Robust LM (error)	1	210.4253	0.00000
Lagrange Multiplier (SARMA)	2	3796.2750	0.00000

Todas as regressões feitas manifestaram heteroescasticidade. A introdução das variáveis de distância é irrelevante porque nenhuma delas é estatisticamente significativa como variável explicativa na regressão OLS.

Tendo passado à GWR, os resultados dão um modelo onde só as variáveis altitude média e declive médio são as variáveis explicativas. À semelhança do *cluster* 1, os coeficientes também têm uma distribuição em curvas de nível no espaço analisado, patente nas Figura 16 e Figura 17.

Figura 16 - Coeficientes da variável Altitude Média por polígono da regressão
GWRT1A1S1R0L0P1T0CL2

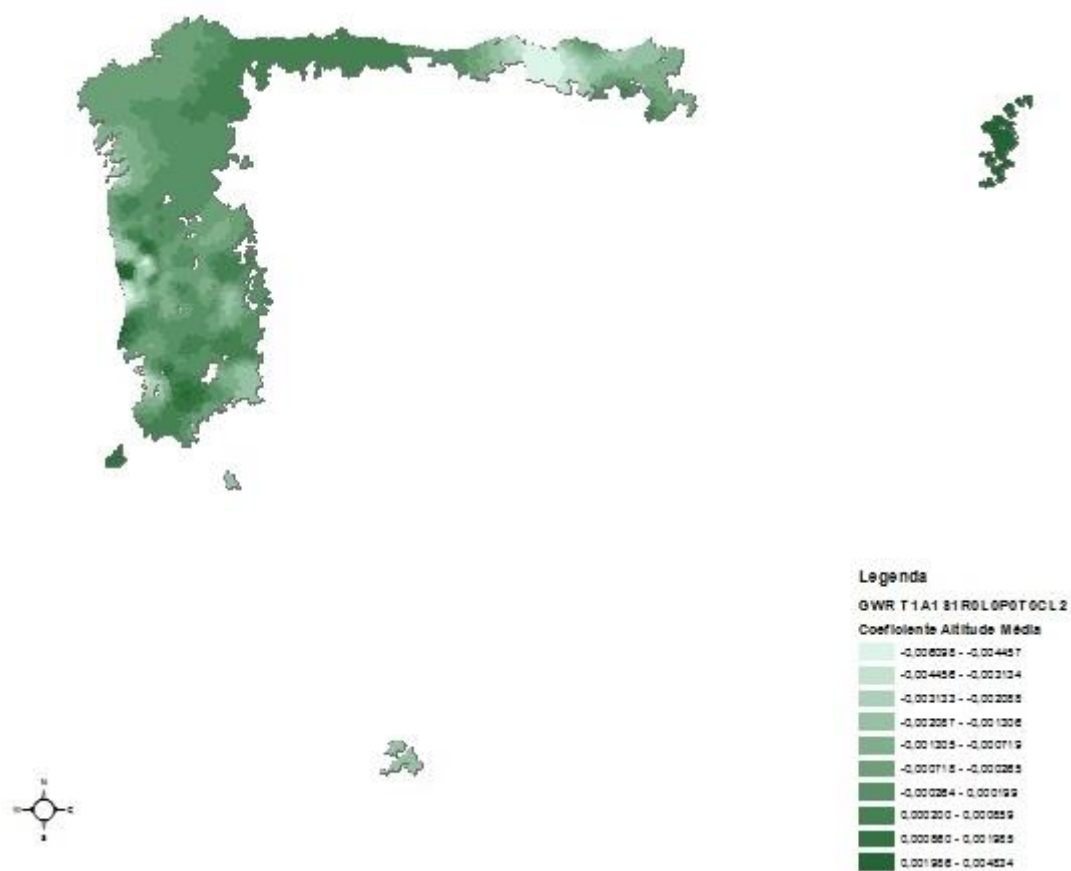
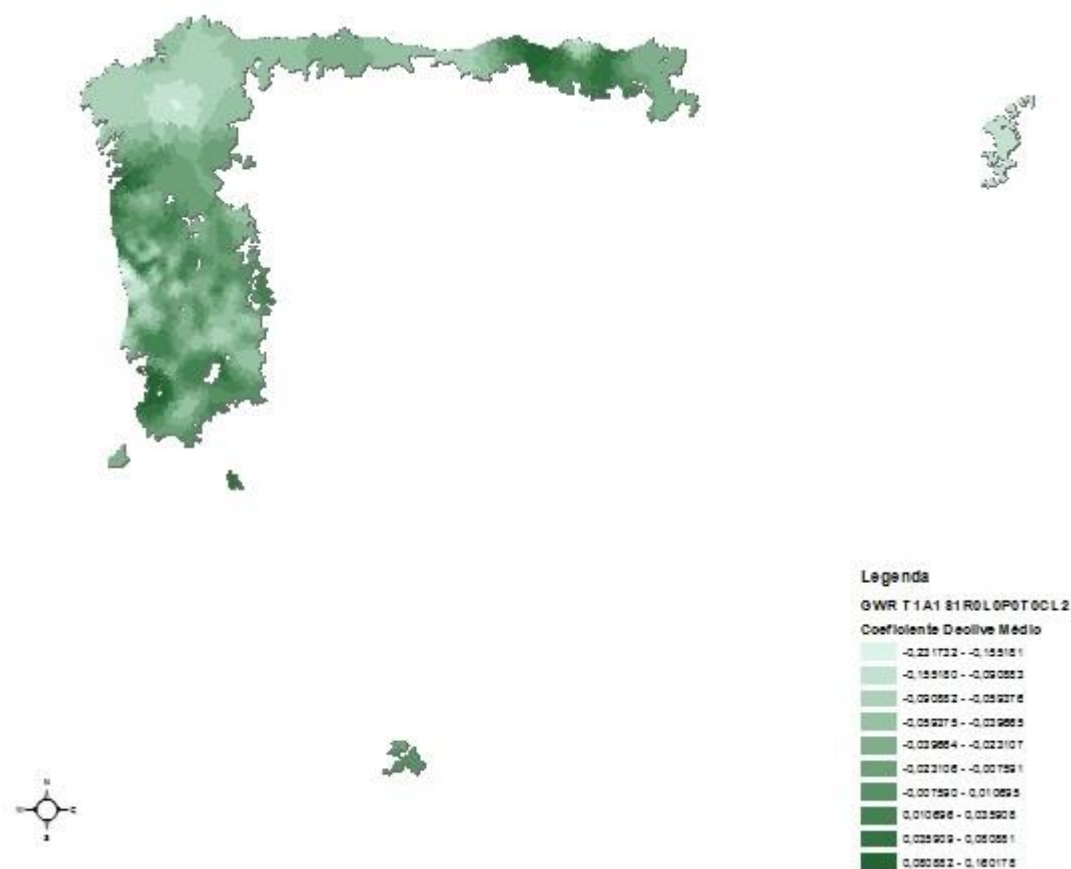
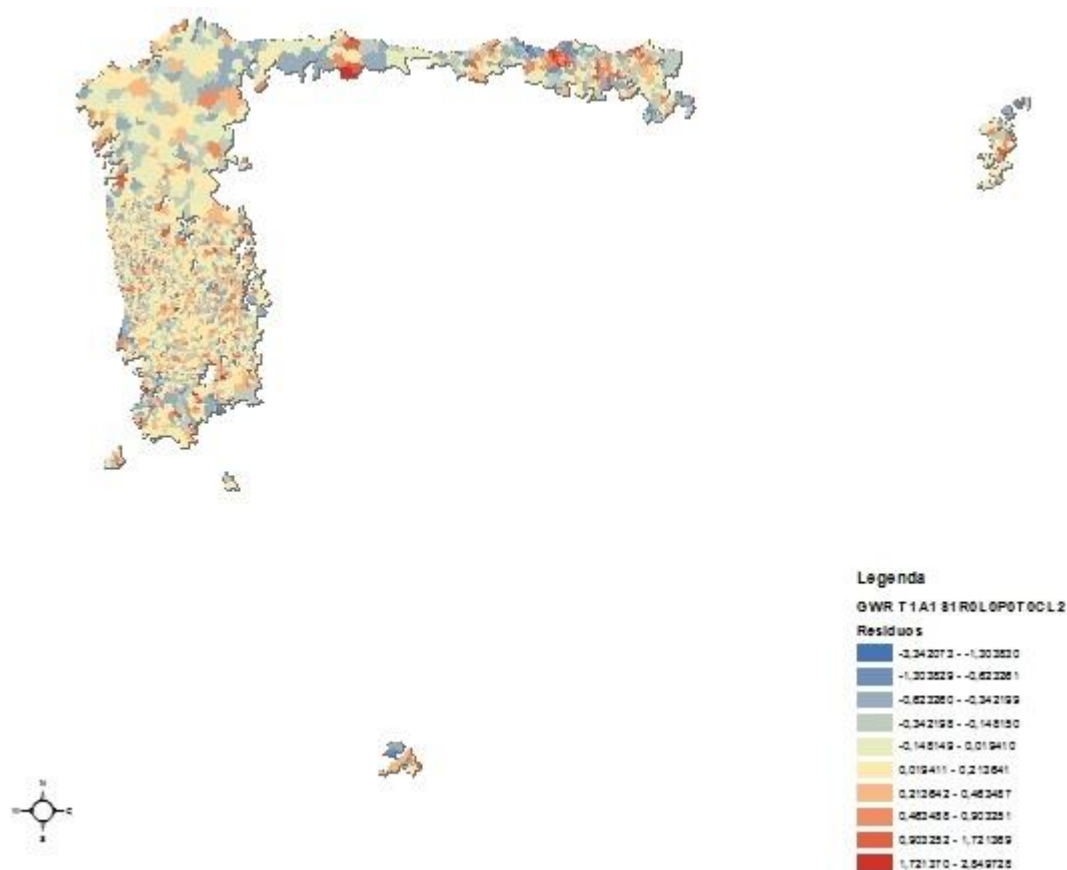


Figura 17 - Coeficientes da variável Declive Médio por polígono da regressão
GWR T1A1 S1R0L0P1T0CL2



Os resíduos também são heterogêneos (vide Figura 18).

Figura 18 - Resíduos da regressão GWRT1A1S1R0L0P1T0CL2



O critério AIC é mais baixo no primeiro período que no segundo como seria de esperar, de acordo com a hipótese da perda de influência das variáveis geográficas na distribuição da população na Península Ibérica

5.2.1.3. Cluster 3

Neste ponto, chegamos à conclusão da necessidade de optar pelo uso da GWR. No entanto, o caminho seguido foi diferente. Se, no caso de quatro variáveis explicativas, todos os testes de dependência espacial são significantes como se constata na Tabela 19, acrescentando as variáveis distância temos a

Tabela 20 já só com o teste significativo de *spatial error model*.

Tabela 19 - Diagnóstico de Dependência Espacial da Regressão T1A1S1L0R0P1T1CL3OLS

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.3352	58.4919	0.00000
Lagrange Multiplier (lag)	1	3141.6216	0.00000
Robust LM (lag)	1	5.7037	0.01693
Lagrange Multiplier (error)	1	3345.6099	0.00000
Robust LM (error)	1	209.6921	0.00000
Lagrange Multiplier (SARMA)	2	3351.3137	0.00000

Tabela 20 - Diagnóstico de Dependência Espacial da Regressão T1A1S1L1R1P1T1CL3OLS

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.3136	54.9882	0.00000
Lagrange Multiplier (lag)	1	2749.1931	0.00000
Robust LM (lag)	1	0.3379	0.56102
Lagrange Multiplier (error)	1	2927.8875	0.00000
Robust LM (error)	1	179.0323	0.00000
Lagrange Multiplier (SARMA)	2	2928.2254	0.00000

O *spatial error model* tem a precipitação como não significativa. A sua retirada ainda deixa heteroscedasticidade presente.

Tendo passado à GWR, os resultados foram negativos ao agrupar todas as variáveis. Não executámos qualquer modelo com mais do que uma variável.

5.2.1.4. Cluster 4

Neste *cluster* podemos observar, através dos testes robustos patentes na Tabela 21, a necessidade de executar um *Spatial Error Model*.

Tabela 21 - Diagnóstico de Dependência Espacial da Regressão T1A1S1L0R0P1T1CL4OLS

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.315	24.7776	0.00000
Lagrange Multiplier	1	568.8771	0.00000
Robust LM (lag)	1	0.3926	0.53093
Lagrange Multiplier	1	584.2095	0.00000
Robust LM (error)	1	15.7251	0.00007
Lagrange Multiplier	2	584.6022	0.00000

Fazendo este modelo com as variáveis em causa temos duas variáveis não significativas, a saber: declive médio e precipitação média, como se pode ver na Tabela 22.

Tabela 22 - Coeficientes e Testes da Regressão T1A1S1L0R0P1T1CL4SEM

Variable	Coefficient	Std.Error	z-value	Probability
CONSTANT	-2.06368700000000	0.61323620000	-3.3652400	0.00076
ALT_MEAN	0.0004291326000	0.00017651840	2.4310920	0.01505
SLP_MEAN	-0.0026251360000	0.00444861000	-0.5901024	0.55512
WCPRECMEAN	0.0001768497000	0.00016881610	1.0475880	0.29483
WCTEMPMEAN	0.0147585200000	0.00364286700	4.0513490	0.00005
LAMBDA	0.6311108000000	0.03166685000	19.9297000	0.00000

Excluindo estas variáveis da regressão, todas as variáveis passam no teste de significância T de Student e no teste de Breusch Pagan com um *P value* de 0.09233 a mostrar homocedasticidade dos resíduos ao nível de significância de 10%.

Tabela 23 - Coeficientes e Testes da Regressão T1A1S0L0R0P0T1CL4SEM

Variable	Coefficient	Std.Error	z-value	Probability
CONSTANT	-1.6873510000000	0.46615330000	-3.6197350	0.00029
ALT_MEAN	0.0003587613000	0.00016303550	2.2005100	0.02777
WCTEMPMEAN	0.0127849700000	0.00298447400	4.2838250	0.00002
LAMBDA	0.6344662000000	0.03144265000	20.1785200	0.00000

À exceção da constante, todos os coeficientes são significativamente positivos.

O coeficiente de determinação (R²) passa de 0,083475, do modelo OLS, para 0,315312, o modelo com as variáveis significantes. Comparando os três modelos executados para este período e para este *cluster*, temos uma diminuição expectável do critério AIC de 1754,65 para 1456,94. Consideramos este modelo bom.

5.2.2. Período 1940-2001

5.2.2.1. Cluster 1

Neste *cluster* podemos observar, através dos testes robustos constantes da Tabela 24, a necessidade de executar um *Spatial Error Model*.

Tabela 24 - Diagnóstico de Dependência Espacial da Regressão T2A1S1L0R0P1T1CL1OLS

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.5756	89.6265	0.00000
Lagrange Multiplier (lag)	1	7100.5874	0.00000
Robust LM (lag)	1	0.0330	0.85590
Lagrange Multiplier (error)	1	7849.7849	0.00000
Robust LM (error)	1	749.2304	0.00000
Lagrange Multiplier (SARMA)	2	7849.8178	0.00000

Seguindo este modelo com as variáveis em causa temos duas variáveis não significativas, a saber: temperatura média e precipitação média – vide Tabela 25.

Tabela 25 - Coeficientes e Testes da Regressão T2A1S1L0R0P1T1CL1SEM

Variable	Coefficient	Std.Error	z-value	Probability
CONSTANT	0.0424779500000	2.32268300000	0.0182883	0.98541
ALT_MEAN	-0.0024692090000	0.00064159810	-3.8485290	0.00012
SLP_MEAN	-0.0449082500000	0.01185503000	-3.7881180	0.00015
WCPRECMEAN	0.0005451088000	0.00072459070	0.7522989	0.45187
WCTEMPMEAN	0.0023164480000	0.01256122000	0.1844127	0.85369
LAMBDA	0.8920040000000	0.01016424000	87.7590400	0.00000

Excluindo estas variáveis da regressão, todas as variáveis passam no teste de significância T de Student, conforme a Tabela 26, e com o teste de Breusch Pagan (*P value* nulo) a mostrar heteroscedasticidade. Incluindo as variáveis Distância ao Litoral e Distância aos Rios Principais, também temos um modelo heteroscedástico.

Tabela 26 - Coeficientes e Testes da Regressão T2A1S1L0R0P0T0CL1SEM

Variable	Coefficient	Std.Error	z-value	Probability
CONSTANT	0.7427673000000	0.15170330000	4.8961840	0.00000
ALT_MEAN	-0.0025051440000	0.00025934700	-9.6594300	0.00000
SLP_MEAN	-0.0441268200000	0.01181230000	-3.7356650	0.00019
LAMBDA	0.8914164000000	0.01020699000	87.3339200	0.00000

Prosseguimos com a execução da GWR.

Na GWR há um erro de especificação no modelo com as quatro variáveis independentes. Fazendo GWR para cada uma das variáveis só não se consegue obter resultados com a Temperatura Média. Fazendo uma regressão com as três restantes variáveis, obtém-se uma regressão com os resultados constantes na Figura 19, Figura 20 e Figura 21. Em cada uma das figuras podemos ver os coeficientes polígono a polígono. De notar que há uma continuidade dos valores nas freguesias contíguas para os coeficientes das três variáveis independentes.

Figura 19 - Coeficientes da variável Altitude Média por polígono da regressão
GWRT2A1S1R0L0P1T0CL1

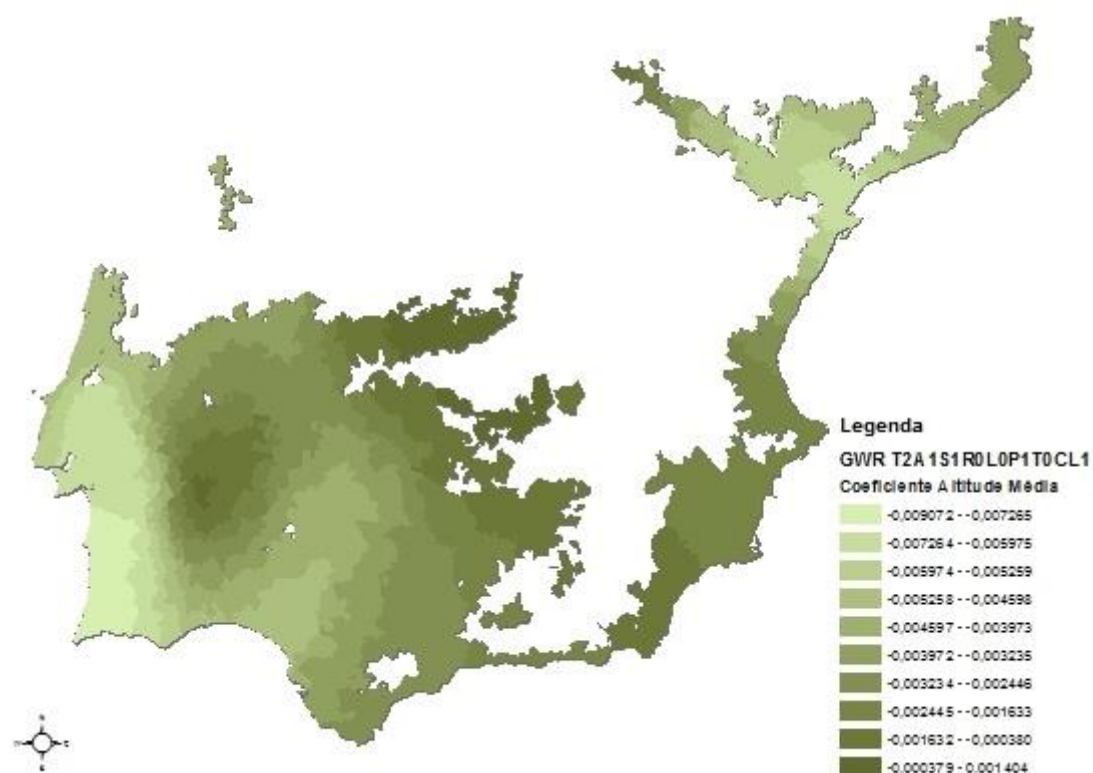


Figura 20 - Coeficientes da variável Declive Médio por polígono da regressão
GWRT2A1S1R0L0P1T0CL1

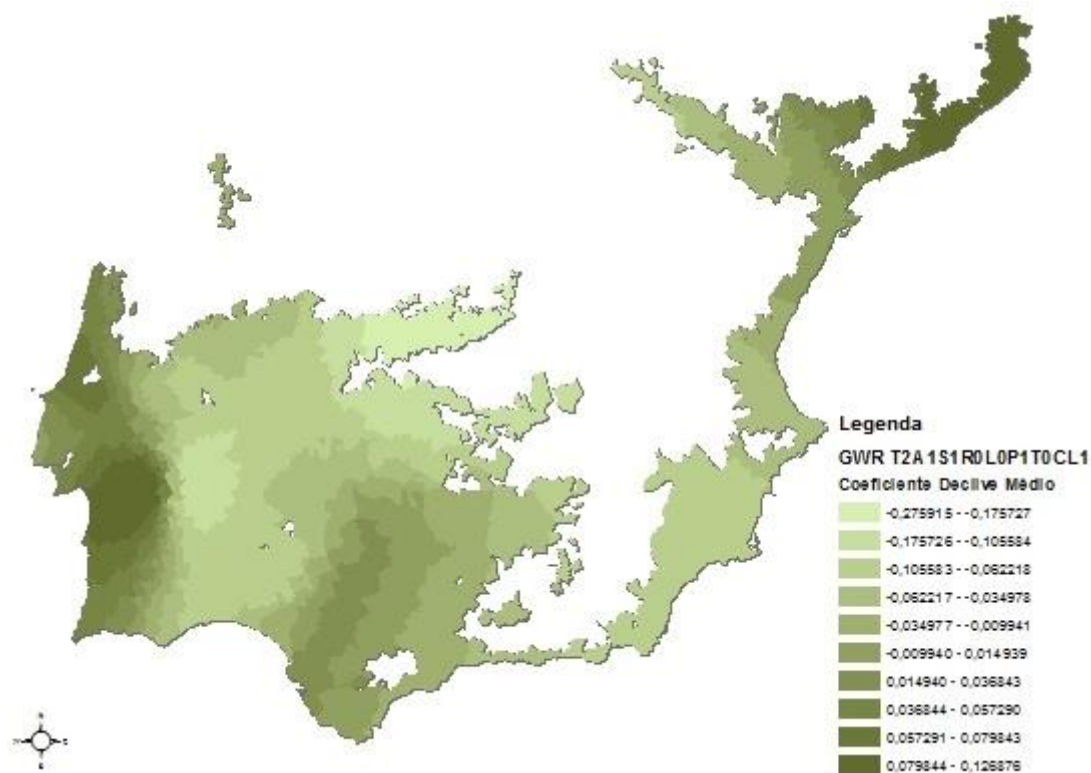
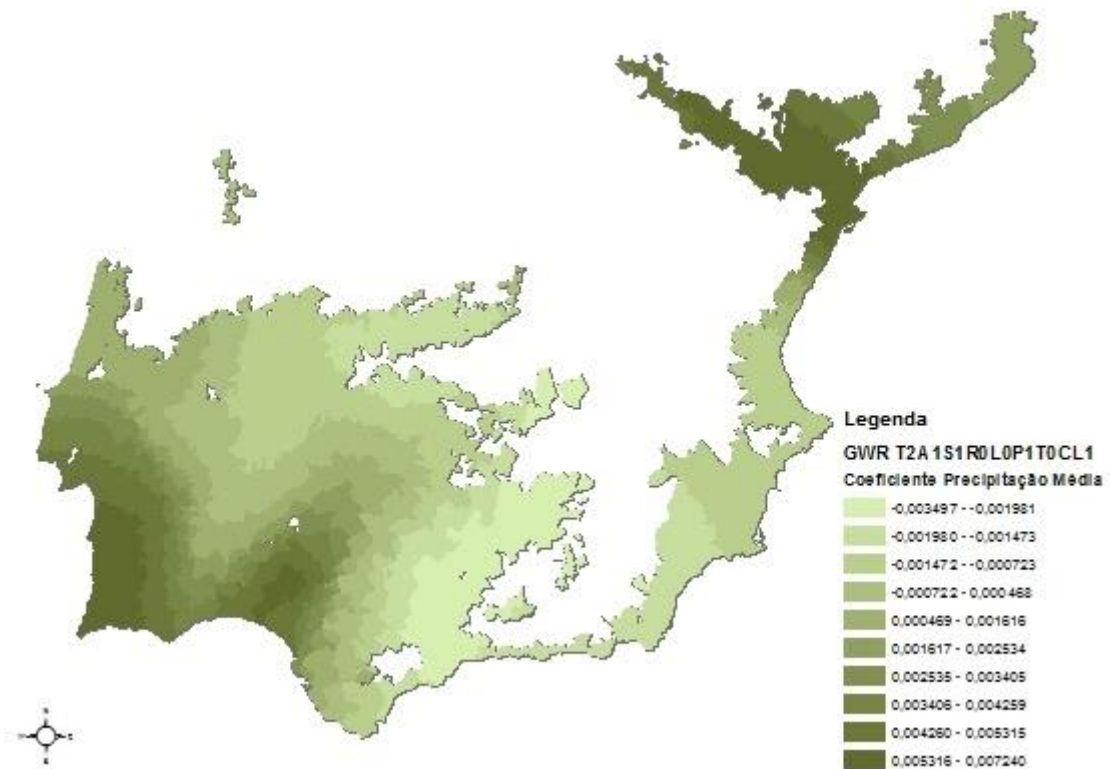


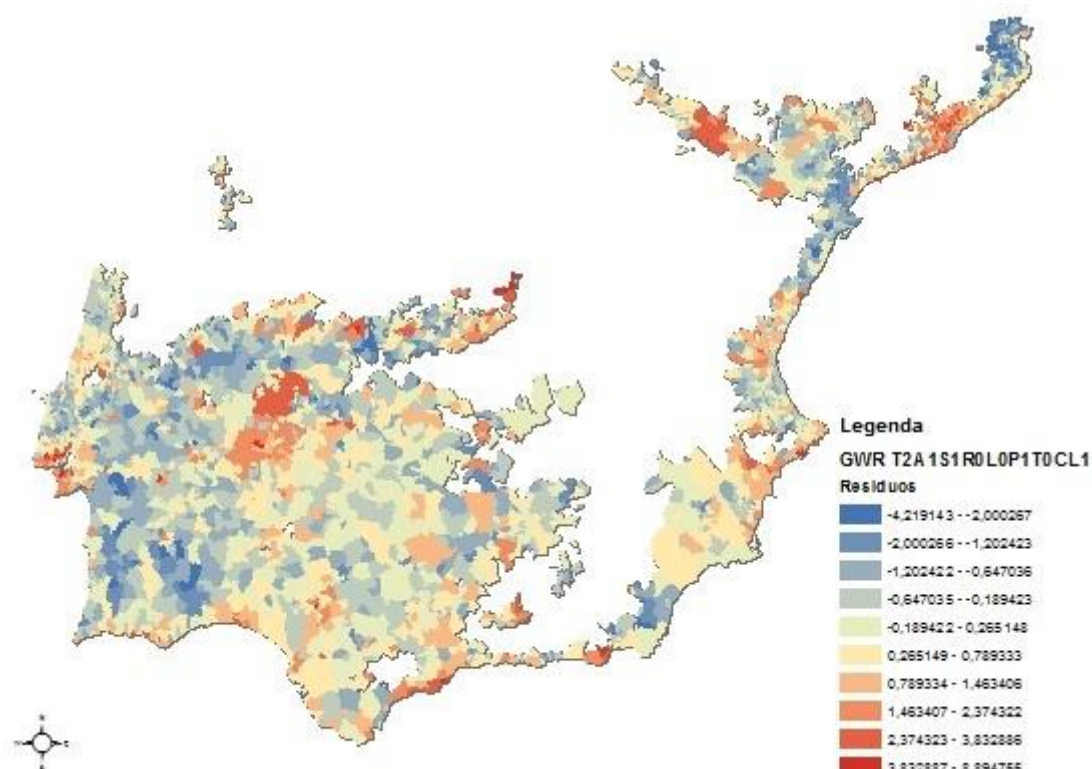
Figura 21 - Coeficientes da variável Precipitação Média por polígono da regressão
GWRT2A1S1R0L0P1T0CL1



O AIC sobe em relação aos melhores modelos obtidos em SEM e SLM.

Os Resíduos não têm uma distribuição homogênea como é traduzido na Figura 22. Este é um sintoma da inexistência de autocorrelação espacial.

Figura 22 - Resíduos da regressão GWRT2A1S1R0L0P1T0CL1



5.2.2.2. Cluster 2

Neste *cluster* podemos observar, através dos testes robustos, a necessidade de executar um *Spatial Error Model* incluindo as variáveis Distância ao Litoral e Distância aos Rios Principais. Se, sem estas variáveis, não encontramos um caminho a seguir, com elas o *Spatial Error Model* parece ser o mais indicado conforme se pode ver nas Tabela 27 e Tabela 28.

Tabela 27 - Diagnóstico de Dependência Espacial da Regressão T2A1S1L0R0P1T1CL2OLS

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.4373	71.8670	0.00000
Lagrange Multiplier (lag)	1	3981.3498	0.00000
Robust LM (lag)	1	6.2649	0.01232
Lagrange Multiplier (error)	1	5036.7635	0.00000
Robust LM (error)	1	1061.6786	0.00000
Lagrange Multiplier (SARMA)	2	5043.0284	0.00000

Tabela 28 - Diagnóstico de Dependência Espacial da Regressão T2A1S1L1R1P1T1CL2OLS

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.4238	69.9927	0.00000
Lagrange Multiplier (lag)	1	3751.0044	0.00000
Robust LM (lag)	1	1.0791	0.29889
Lagrange Multiplier (error)	1	4729.1314	0.00000
Robust LM (error)	1	979.2061	0.00000
Lagrange Multiplier (SARMA)	2	4730.2105	0.00000

Fazendo este modelo pode-se ver que as variáveis Distância ao Litoral, Distância aos Rios Principais e Temperatura Média não são significativos, conforme a Tabela 29.

Tabela 29 - Coeficientes e Testes da Regressão T2A1S1L1R1P1T1CL2SEM

Variable	Coefficient	Std.Error	z-value	Probability
CONSTANT	-1.9335610000000	1.42618600000	-1.3557570	0.17518
ALT_MEAN	-0.0015430890000	0.00046795470	-3.2975170	0.00098
SLP_MEAN	-0.0711593800000	0.00588723700	-12.0870600	0.00000
DIST_LITOR	-0.0000043521980	0.00000239102	-1.8202240	0.06872
DIST_RIOSP	-0.0000047288340	0.00000425496	-1.1113690	0.26641
WCPRECMEAN	0.0017662620000	0.00038070040	4.6395070	0.00000
WCTEMPMEAN	0.0072878620000	0.00858383700	0.8490215	0.39587
LAMBDA	0.8362352000000	0.01527293000	54.7527900	0.00000

Excluindo estas três variáveis e fazendo uma nova regressão, ficamos com todas as variáveis significativas mas o teste de Breusch Pagan mostra novamente heteroescedasticidade com um *P value* nulo, conforme Tabela 30.

Tabela 30 - Coeficientes e Testes da Regressão T2A1S1L0R0P1T0CL2SEM

Variable	Coefficient	Std.Error	z-value	Probability
CONSTANT	-1.4932000000000	0.31311010000	-4.7689290	0.00000
ALT_MEAN	-0.0022048820000	0.00015970480	-13.8059800	0.00000
SLP_MEAN	-0.0695004100000	0.00585339500	-11.8735200	0.00000
WCPRECMEAN	0.0022104600000	0.00029556360	7.4787970	0.00000
LAMBDA	0.8403659000000	0.01499830000	56.0307300	0.00000

Tendo passado à GWR, e à semelhança do período 1878-1940, os resultados dão origem a um modelo onde só as variáveis altitude média e declive médio são as variáveis explicativas. À semelhança do *cluster* 1, os coeficientes também têm uma distribuição em curvas de nível no espaço analisado, patente nas Figura 23 e Figura 24.

Figura 23 - Coeficientes da variável Altitude Média por polígono da regressão
GWRT2A1S1R0L0P0T0CL2

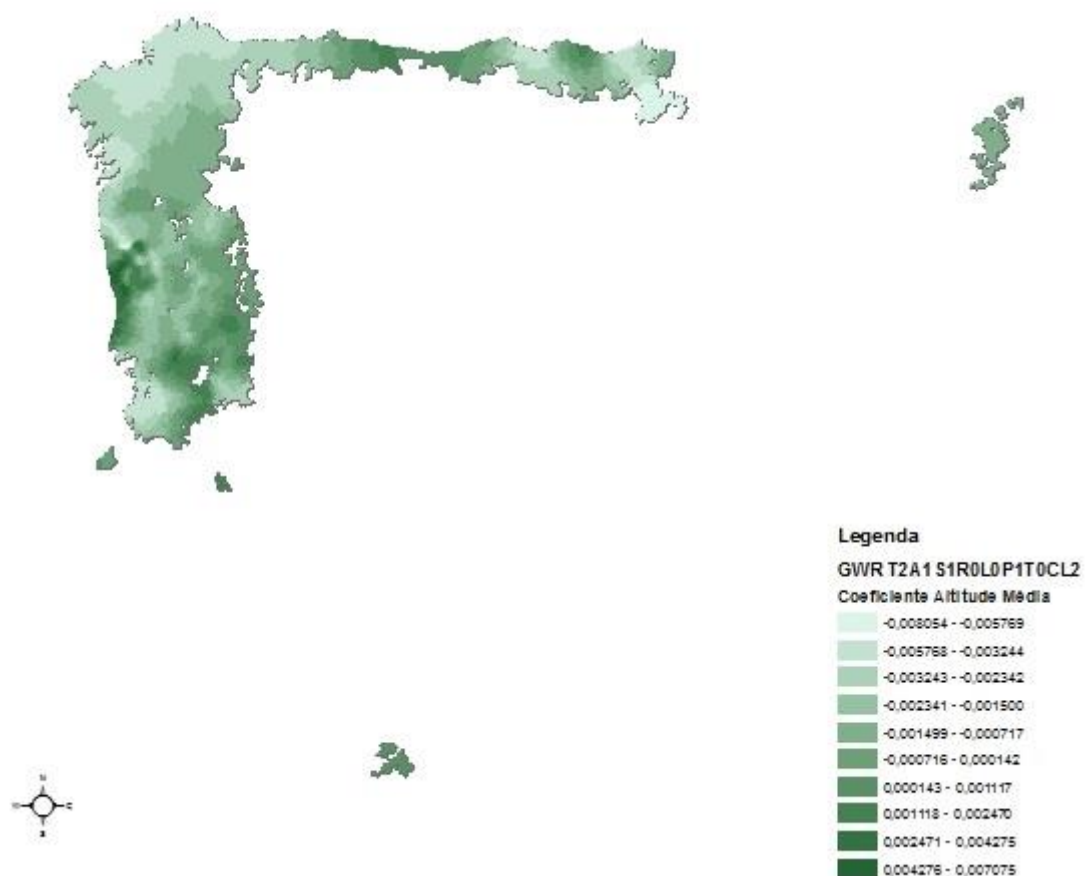
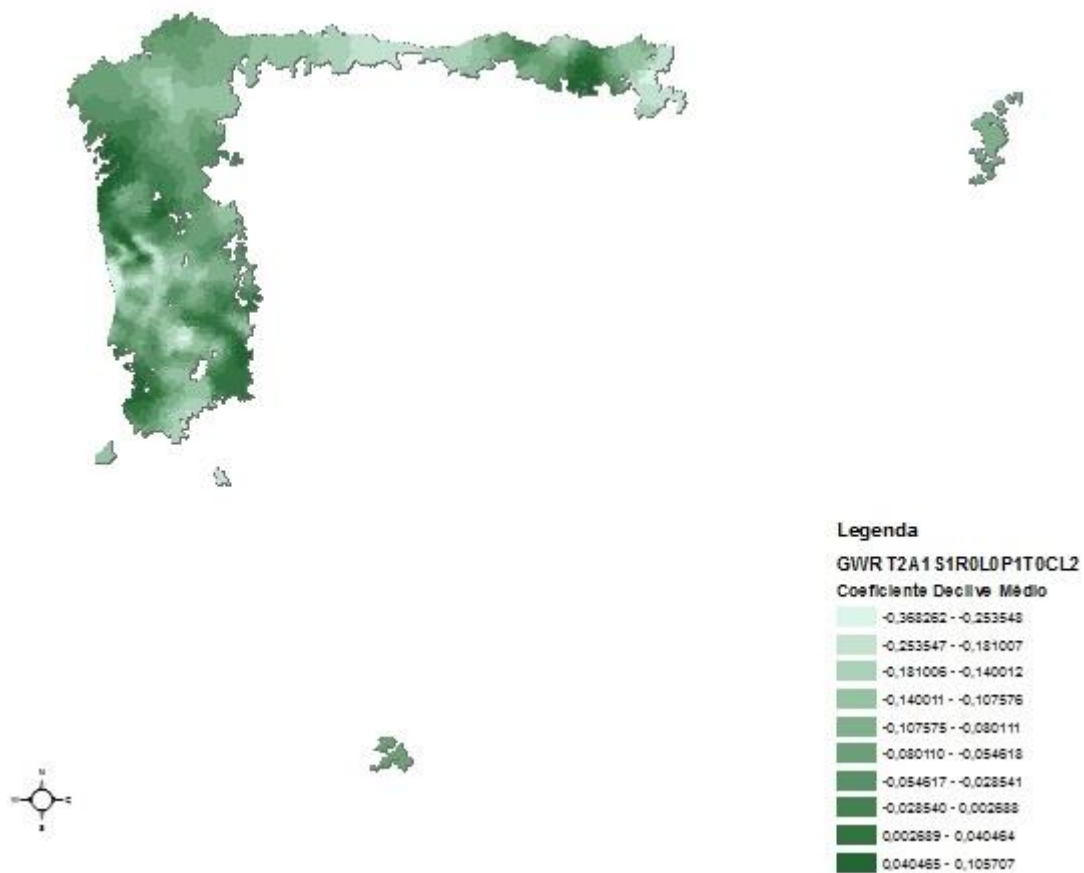
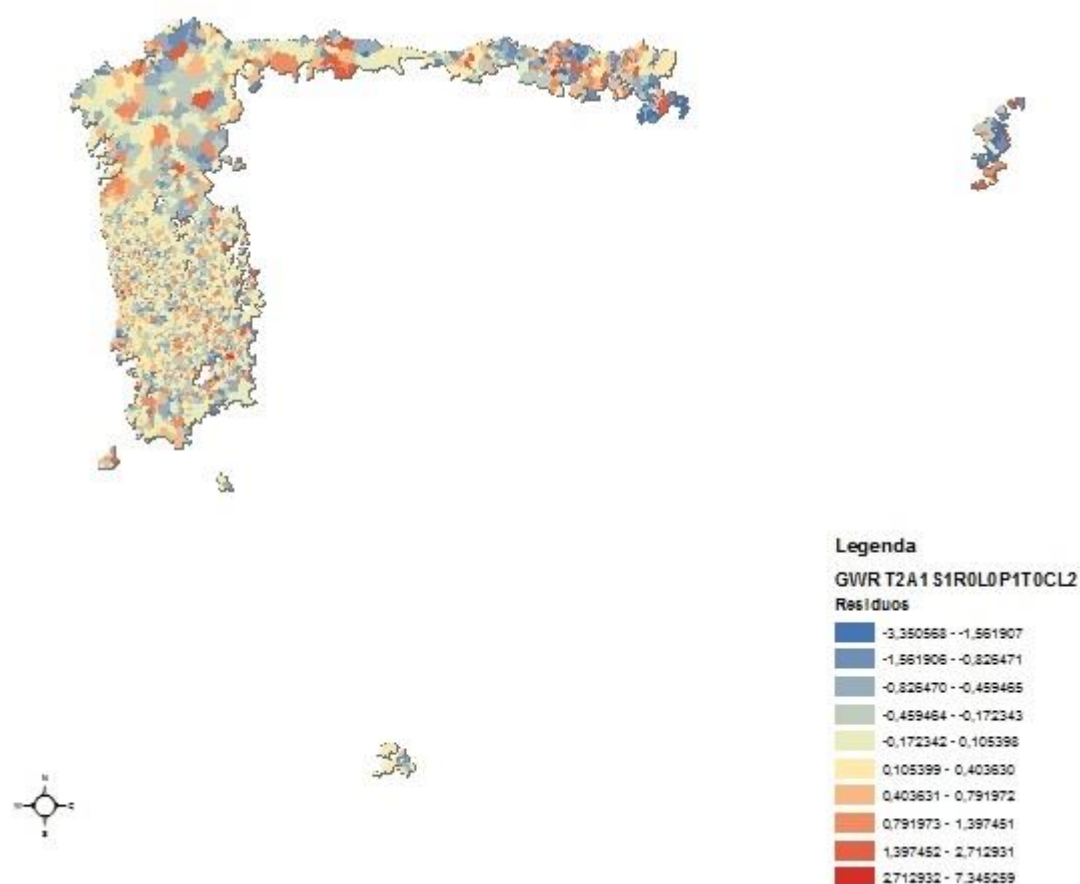


Figura 24 - Coeficientes da variável Declive Médio por polígono da regressão
GWRT2A1S1R0L0P0T0CL2



Uma vez mais, a heterogeneidade do mapa dos resíduos (Figura 25) demonstra a inexistência de autocorrelação espacial.

Figura 25 - Resíduos da regressão GWRT2A1S1R0L0P0T0CL2



5.2.2.3. Cluster 3

Neste *cluster*, os 4 testes de dependência espacial não são esclarecedores para a escolha entre o *Spatial Lag Model* ou para o *Spatial Error Model*, incluindo ou não na regressão as variáveis Distância ao Litoral e Distância aos Rios Principais, como se pode ver nas Tabela 31 e Tabela 32.

Tabela 31 - Diagnóstico de Dependência Espacial da Regressão T2A1S1L0R0P1T1CL3OLS

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.5135	89.4880	0.00000
Lagrange Multiplier (lag)	1	6954.7890	0.00000
Robust LM (lag)	1	39.1693	0.00000
Lagrange Multiplier (error)	1	7848.5376	0.00000
Robust LM (error)	1	932.9179	0.00000
Lagrange Multiplier (SARMA)	2	7887.7069	0.00000

Tabela 32 - Diagnóstico de Dependência Espacial da Regressão T2A1S1L1R1P1T1CL3OLS

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.4959	86.7859	0.00000
Lagrange Multiplier (lag)	1	6518.2403	0.00000
Robust LM (lag)	1	37.5477	0.00000
Lagrange Multiplier (error)	1	7319.6918	0.00000
Robust LM (error)	1	838.9991	0.00000
Lagrange Multiplier (SARMA)	2	7357.2395	0.00000

Fazendo *Spatial Error Model* e *Spatial Lag Model* obtemos um modelo de cada - Tabela 33 e Tabela 34, respetivamente – com variáveis significativas, mas com presença de heteroscedasticidade.

Tabela 33 - Coeficientes e Testes da Regressão T2A0S1L1R1P1T1CL3SEM

Variable	Coefficient	Std.Error	z-value	Probability
CONSTANT	-8.6076940000000	0.95555460000	-9.0080610	0.00000
SLP_MEAN	-0.0617884800000	0.00928145200	-6.6571990	0.00000
DIST_LITOR	0.0000029091300	0.00000130073	2.2365450	0.02532
DIST_RIOSP	-0.0000053895890	0.00000232356	-2.3195380	0.02037
WCPRECMEAN	0.0027296880000	0.00070766580	3.8573120	0.00011
WCTEMPMEAN	0.0456437700000	0.00442340300	10.3187000	0.00000
LAMBDA	0.8666776000000	0.01148861000	75.4379600	0.00000

Tabela 34 - Coeficientes e Testes da Regressão T2A1S1L0R0P0T1CL3SLM

Variable	Coefficient	Std.Error	z-value	Probability
W_TCAM194001	0.8420741000000	0.01204497000	69.9108800	0.00000
CONSTANT	-1.3844800000000	0.24091150000	-5.7468400	0.00000
ALT_MEAN	-0.0003191555000	0.00009546010	-3.3433400	0.00083
SLP_MEAN	-0.0324258600000	0.00485910900	-6.6732110	0.00000
WCTEMPMEAN	0.0124528100000	0.00161207500	7.7247120	0.00000

Neste *cluster*, à semelhança do que acontecia no período pré-Segunda Guerra Mundial, a execução da *Geographic Weighted Regression* não resultou em qualquer modelo com pluralidade de variáveis explicativas.

5.2.2.4. Cluster 4

Neste *cluster* podemos observar na Tabela 35, através dos testes robustos, a presença de um *Spatial Lag Model*.

Tabela 35 - Diagnóstico de Dependência Espacial da Regressão T2A1S1L0R0P1T1CL4OLS

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.3296	25.9206	0.00000
Lagrange Multiplier (lag)	1	670.7051	0.00000
Robust LM (lag)	1	31.2561	0.00000
Lagrange Multiplier (error)	1	639.9424	0.00000
Robust LM (error)	1	0.4933	0.48244
Lagrange Multiplier (SARMA)	2	671.1984	0.00000

Fazendo este modelo com as variáveis em causa temos a altitude média não significativa (Tabela 36).

Tabela 36 - Coeficientes e Testes da Regressão T2A1S1L0R0P1T1CL4SLM

Variable	Coefficient	Std.Error	z-value	Probability
W_TCAM194001	0.6789642000000	0.02818621000	24.0885300	0.00000
CONSTANT	-2.5411240000000	0.53823420000	-4.7212240	0.00000
ALT_MEAN	0.0002668519000	0.00015259250	1.7487880	0.08033
SLP_MEAN	0.0163679300000	0.00624810700	2.6196620	0.00880
WCPRECMEAN	0.0005460960000	0.00015780880	3.4604910	0.00054
WCTEMPMEAN	0.0109527300000	0.00310222900	3.5306010	0.00041

Ao excluir esta variável da regressão, todas as variáveis passam no teste de significância (Tabela 37) e no teste de Breusch Pagan com um *P value* de 0.93549 a mostrar homoescedasticidade.

Tabela 37 - Coeficientes e Testes da Regressão T2A0S1L0R0P1T1CL4SLM

Variable	Coefficient	Std.Error	z-value	Probability
W_TCAM194001	0.6876950000000	0.02753357000	24.9766000	0.00000
CONSTANT	-1.7359390000000	0.26314290000	-6.5969470	0.00000
SLP_MEAN	0.0178893900000	0.00617034800	2.8992520	0.00374
WCPRECMEAN	0.0004175086000	0.00013886000	3.0066860	0.00264
WCTEMPMEAN	0.0068121890000	0.00191450300	3.5582030	0.00037

O R2 passa de 0,133667, no modelo OLS, para 0,392094, o modelo com as variáveis significativas. Consideramos este modelo bom.

5.2.3. Comparação entre os Dois Períodos

Observando só a parte dos modelos *Ordinary Least Squares*, *Spatial Error Model* e *Spatial Lag Model*, em ambos os períodos e para qualquer *cluster*, temos coeficientes para as diferentes variáveis com valores quase insignificantes. São valores muito próximos de 0. No caso dos parâmetros de erro espacial e autorregressivo temos valores analisáveis.

Os *clusters* 1 e 2 apresentam heterogeneidade de resultados dentro dos dois períodos estudados. Apesar de o seu limitado poder explicativo das TCAM, este ainda existe, ainda que em pequenos grupos de dados selecionados pelo GWR;

O *cluster* 3 tem uma variação extrema das Taxas de Crescimento Anual Médio muito acentuada entre os dois períodos. Com valores tão díspares nas TCAM e com as variáveis geográficas a manterem-se, deduz-se daqui que a mobilidade das populações se deveu a outros fatores que não os geográficos em análise.

O *cluster* 4 é o mais inóspito dos *clusters* gerados por este trabalho – é o que melhores resultados apresenta com a aderência aos seus dados de um *Spatial Error Model*, no período 1878-1940, e um *Spatial Lag Model*, no período 1940-2001. Estes resultados não são de espantar, visto que é o *cluster* mais despovoado. O fator crítico a analisar é a passagem de um Modelo de Erro Espacial para um Modelo de Desfasamento Espacial, do primeiro para o segundo período. Podemos dizer que os efeitos das variáveis se passam a fazer sentir nos polígonos vizinhos. O efeito dos erros mantém-se.

Apesar de só no caso do *cluster* 4 se tirar conclusões pela validade dos modelos escolhidos, fizemos vários modelos de Erro ou Desfasamento Espacial. À exceção de um caso, optámos sempre pelo Modelo de Erro Espacial. Isto vai de encontro à perda de importância das variáveis geográficas na Taxa de Crescimento Anual Médio das populações. Caso não fosse assim, o modelo preferencial seria o Modelo de Desfasamento Espacial.

Observando o quadro comparativo de AIC entre períodos, conforme a Tabela 38, os resultados são sempre piores no segundo período. Isto vai de encontro à tese, segundo a qual, o peso dos fatores geográficos na explicação da distribuição espacial da população diminuiu no período de 1940-2001.

Tabela 38 - Quadro comparativo de AIC entre períodos

Variáveis Independentes e Cluster	Diferença entre Variáveis Dependentes T2 - T1
A1S1L0R0T1P1CL1	3 737.14
A1S1L0R0T1P1CL2	3 883.51
A1S1L0R0T1P1CL3	7 501.40
A1S1L0R0T1P1CL4	1 910.25
A1S1L1R1T1P1CL1	3 560.99
A1S1L1R1T1P1CL2	3 777.86
A1S1L1R1T1P1CL3	7 504.80
A1S1L1R1T1P1CL4	1 904.29

6. CONCLUSÕES

O estudo teve por objetivo investigar a hipótese: até início da II guerra (1878-1940) os fatores geográficos explicam grande parte da distribuição espacial da população; no período 1940-2000, foram os fatores socioeconómicos (atração das grandes cidades) que passaram a explicar as alterações nos padrões da população. Os resultados obtidos através das análises de regressão para os dois períodos, trazem novas evidências para suportar esta hipótese.

Deste objectivo principal derivam as seguintes conclusões para os objectivos os quais caracterizam esta problemática:

O número de segmentos (*clusters*) que otimiza o espaço ibérico é de quatro nos dois períodos. Em ambos os casos são bastante heterogéneos e atravessam a fronteira entre os dois estados ibéricos.

O *cluster* 1 é caracterizado por elevadas taxas de crescimento anual médio (TCAM), em ambos os períodos, e por elevadas temperaturas médias. Geograficamente, inclui a maior parte da região sul da Península Ibérica (sul do rio Tejo, Extremadura e parte ocidental da Andaluzia) e o litoral leste de Espanha.

O *cluster* 2 engloba o Norte de Portugal, que faz um contínuo com a Galiza, Astúrias, Cantábria e País Basco. A divisão Norte / Sul é uma divisão que se faz no vale do Rio Tejo, desde a fronteira até aos concelhos do litoral do Distrito de Leiria. É o *cluster* com mais alta pluviosidade.

O *cluster* 3 situa-se quase na sua totalidade no interior da Península Ibérica e em Espanha, e inclui o maior número de polígonos (freguesias/municípios). No segundo período, este *cluster* inclui polígonos com taxas de crescimento anual médio que são, em média, bastante mais baixas do que no primeiro período.

O *cluster* 4 engloba as Serras e Parques Naturais e, em geral, zonas inóspitas, com baixas temperaturas médias e alta pluviosidade, rugosas e com altitude elevada.

Apesar de os segmentos definidos terem demonstrado a continuidade entre os dois lados da fronteira, as mudanças de *cluster* não ocorreram de forma simétrica. Verificou-se que a probabilidade de um polígono espanhol mudar de *cluster* é 1,20 vezes mais provável do que um polígono português. Este resultado poderá ser explicado por uma maior atração dos grandes centros urbanos espanhóis.

Os polígonos que passaram do *cluster* 3 para o 1 refletem a urbanização dos grandes centros urbanos espanhóis de Madrid (com Toledo), Barcelona (com Tarragona e Lleida), Valencia (com Alicante), Navarra e Saragoça. A população sextuplicou de 1940 para 2001 nestas áreas.

Os polígonos que passaram do *cluster* 1 para o 3 (apenas Espanhois) são os que tinham uma TCAM média mais elevada em 1878-1940 (139%), e passaram a ter uma TCAM média negativa de -142%. São municípios maioritariamente das províncias de Ciudad Real, Granada, Jaén e Toledo.

Os polígonos que passam do *cluster* 2 para o 3 são 83% portugueses. Demonstram a desertificação do interior luso nos distritos de Bragança e Guarda, predominando neste grupo freguesias dos concelhos raianos de Almeida e Sabugal.

As variáveis geográficas que explicam o crescimento anual médio da população em cada período da Península Ibérica resumem-se a: altitude média, declive médio, precipitação média e temperatura média. No entanto, nem todas elas são significativamente explicativas em cada *cluster*. Se no *cluster* 1, a altitude, o declive e a precipitação são usadas para explicar a TCAM dos dois períodos através de um Modelo *Geographic Weighted Regression*, já no *cluster* 2 mantém-se o mesmo modelo mas só com a altitude e o declive explicam a evolução da população. O *cluster* 3 mostra-se o mais difícil de trabalhar com as variáveis em causa, porque nenhum dos modelos testados provou a utilidade destas variáveis para explicar as TCAM. O *cluster* 4 foi o que melhores resultados obteve como um todo. Não necessitou de ser fraccionado com uma GWR para se obterem resultados. Um *Spatial Error Model* no período de 1878-1940 com as variáveis altitude e temperatura e um *Spatial Lag Model* no período 1940-2001 com as variáveis declive, precipitação e temperatura foram os modelos que se provaram adequados. A baixa temperatura, o elevado declive e a elevada altitude causam a inospitalidade patenteada em ambos períodos em análise pela pior taxa de crescimento média dos quatro segmentos.

6.1. LIMITAÇÕES E RECOMENDAÇÕES PARA TRABALHOS FUTUROS

As variáveis precipitação média e temperatura média têm dados retirados dos *rasters* “ESRI grids” do projeto WorldClim (<http://www.worldclim.org/current>). Estes dados estavam disponíveis numa base mensal e só existem dados a partir de 1950. Caso existissem dados mais antigos poderíamos fazer uma quebra neste tipo de dados, à semelhança do que foi feito para as variáveis demográficas.

Em vez da utilização da metodologia k means para executar a segmentação dos dados sugere-se a utilização do SOM.

Variáveis sociais e de equipamentos económicos e sociais são inexistentes para todo o horizonte temporal visado por este projecto ao nível das freguesias e municípios. É uma limitação geradora de tristeza na medida em que, caso não existisse, poder-se-ia chegar a resultados com uma maior capacidade explicativa do fenómeno estudado.

7. ANEXOS

Tabela 39 - Lista de Polígonos sem Dados Climatéricos

Número do Polígono	Nome	Província / Distrito	País	Designação Administrativa
1.225	Borba (São Bartolomeu)	Borba	Portugal	Freguesia
1.265	Reboleira	Amadora	Portugal	Freguesia
1.270	Alvalade	Lisboa	Portugal	Freguesia
1.293	Coração de Jesus	Lisboa	Portugal	Freguesia
1.302	Pena	Lisboa	Portugal	Freguesia
1.303	Socorro	Lisboa	Portugal	Freguesia
1.305	Santa Justa	Lisboa	Portugal	Freguesia
1.307	Encarnação	Lisboa	Portugal	Freguesia
1.309	São Vicente de Fora	Lisboa	Portugal	Freguesia
1.312	Castelo	Lisboa	Portugal	Freguesia
1.315	São Miguel	Lisboa	Portugal	Freguesia
1.316	Santiago	Lisboa	Portugal	Freguesia
1.318	São Nicolau	Lisboa	Portugal	Freguesia
1.319	Santa Catarina	Lisboa	Portugal	Freguesia
1.320	Sé	Lisboa	Portugal	Freguesia
1.321	Mártires	Lisboa	Portugal	Freguesia
1.322	São Paulo	Lisboa	Portugal	Freguesia
1.324	Madalena	Lisboa	Portugal	Freguesia
1.351	Barreiro	Barreiro	Portugal	Freguesia
1.389	Setúbal (Santa Maria da Graça)	Setúbal	Portugal	Freguesia
1.629	Évora (São Mamede)	Évora	Portugal	Freguesia
1.631	Évora (Santo Antão)	Évora	Portugal	Freguesia
1.787	Lugar Nuevo de la Corona	Valencia	Espanha	Município
2.522	Peniche (São Pedro)	Peniche	Portugal	Freguesia
2.862	Coimbra (São Bartolomeu)	Coimbra	Portugal	Freguesia
3.180	Vitória	Porto	Portugal	Freguesia
3.184	São Nicolau	Porto	Portugal	Freguesia
3.575	Trandeiras	Braga	Portugal	Freguesia
3.695	Guimarães (São Sebastião)	Guimarães	Portugal	Freguesia
7.212	Caminha (Matriz)	Caminha	Portugal	Freguesia
9.566	Sant Hipòlit de Voltregà	Barcelona	Espanha	Município
9.921	Maleján	Zaragoza	Espanha	Município
10.187	Badia del Vallès	Barcelona	Espanha	Município
12.148	Comunidad de Canicosa de la Sierra y Vilviestre del Pinar	Burgos	Espanha	Comunidad
12.154	Comunidad de Canicosa de la Sierra y Casarejos	Burgos	Espanha	Comunidad
12.163	Jurisdicción Macomunada de El Boalo y Manzanares el Rea (El Chap	Madrid	Espanha	Jurisdicción

12.170	Comunidad de Quintanaortuño y Celadilla-Sotobrín (Merindad d	Burgos	Espanha	Comunidad
12.171	Comunidad de Quintanaortuño y Villanueva de Río Ubierna	Burgos	Espanha	Comunidad
12.172	Comunidad de Quintanaortuño, Sotopalacios y Villanueva	Burgos	Espanha	Comunidad
12.174	Jurisdicción Mancomunada de Bascuñana y Vitoria de Rioja	Burgos	Espanha	Mancomunidad
12.175	Jurisdicción Mancomunada de Bascuñana y Vitoria de Rioja	Burgos	Espanha	Mancomunidad
12.177	Comunidad de Soto de la Vega y Villazala	León	Espanha	Comunidad
12.183	Comunidad de Quintanalara	Burgos	Espanha	Comunidad
12.185	Comunidad de Los Ausines, Quintanalara y Revilla del Campo	Burgos	Espanha	Comunidad
12.189	Comunidad de Barbadillo de Herreros, Valderrubio y Villajime	Burgos	Espanha	Comunidad
12.203	Comunidad de Quintanar de la Sierra y Vilviestre del Pinar	Burgos	Espanha	Comunidad
12.209	Facería 42	Navarra	Espanha	Territorio Histórico
12.212	Facería 41	Navarra	Espanha	Territorio Histórico
12.217	Facería 103	Navarra	Espanha	Territorio Histórico
12.220	Facería 37	Navarra	Espanha	Territorio Histórico
12.222	Facería 21	Navarra	Espanha	Territorio Histórico
12.232	Facería de Suarbe	Navarra	Espanha	Territorio Histórico
12.233	Facería 16	Navarra	Espanha	Territorio Histórico
12.236	Facería 2	Navarra	Espanha	Territorio Histórico
12.237	Facero Aldape	Navarra	Espanha	Territorio Histórico
12.238	Facería 63	Navarra	Espanha	Territorio Histórico
12.239	Facería 52	Navarra	Espanha	Territorio Histórico
12.247	Facería 105	Navarra	Espanha	Territorio Histórico
12.251	Facería 50	Navarra	Espanha	Territorio Histórico
12.253	Facería 56	Navarra	Espanha	Territorio Histórico
12.254	Facería 55	Navarra	Espanha	Territorio Histórico
12.256	Facería 43	Navarra	Espanha	Territorio Histórico
12.267	Facería 71	Navarra	Espanha	Territorio Histórico
12.268	Facería 65	Navarra	Espanha	Territorio Histórico
12.275	Facería de Lerga y Ujué	Navarra	Espanha	Territorio Histórico
12.279	Jurisdicción de Segura de los Baños y Vivel del Río Martín	Teruel	Espanha	Jurisdicción
12.286	Comunidad de Castrillo de la Reina, Hacinas y Salas de los I	Burgos	Espanha	Comunidad

Tabela 40 - Lista de Polígonos sem População

Número do Polígono	Nome	Província / Distrito	Designação Administrativa
12.146	Comunidad de Benatae y Torres de Albánchez	Jaén	Comunidad
12.147	Comunidad de Cotillas y Villaverde de Guadalimar	Albacete	Comunidad
12.148	Comunidad de Canicosa de la Sierra y Vilviestre del Pinar	Burgos	Comunidad
12.149	Comunidad de Palacios de la Sierra y Vilviestre del Pinar	Burgos	Comunidad
12.150	Comunidad de Cilleruelo de Arriba y Pineda-Trasmonte	Burgos	Comunidad

12.151	Comunidad de Hontoria del Pinar y Palacios de la Sierra	Burgos	Comunidad
12.152	Comunidad de Bahabín de Esgueva, Cilleruelo de Abajo y Santi	Burgos	Comunidad
12.153	Comunidad de Arauzo de Miel y Huerta del Rey	Burgos	Comunidad
12.154	Comunidad de Canicosa de la Sierra y Casarejos	Burgos	Comunidad
12.155	Comunidad de Palacios de la Sierra, San Leonardo de Yag³e (Soria	Burgos	Comunidad
12.156	Comunidad de Quintanarraya y Huerta del Rey	Burgos	Comunidad
12.157	Comunidad de Quintanarraya y Espeja de San Marcelino	Burgos	Comunidad
12.158	Comunidad de Caleruega y Coru±a del Conde	Burgos	Comunidad
12.159	Comunidad de Herrera de Soria, Navafr³ya de Ucero y Ucer	Soria	Comunidad
12.160	Comunidad de Aldehorno	Burgos	Comunidad
12.161	Com-n de Riaza y Sep-lveda	Segovia	Territorio Hist³rico
12.162	Coto Mancomunado de Pedrosillo el Ralo y Villaverde de Guare	Salamanca	Mancomunidad
12.163	Jurisdicci³n Macomunada de El Boalo y Manzanares el Rea (El Chap	Madrid	Jurisdicción
12.164	Comunidad de Albalate de Zorita e Illana	Guadalajara	Comunidad
12.165	Comunidad de Campoo-Cabu³rniga (Hermandad de Campoo de Suso	Cantabria	Comunidad
12.166	Mancomunidad de Nueve Villas de Sopellano	Burgos	Mancomunidad
12.167	Mancomunidad de Quintana del Castillo Y Villagat³n	Le³n	Mancomunidad
12.168	Comunidad de Cubo de Bureba y Santa Mar³ya-Ribarredonda	Burgos	Comunidad
12.169	Mancomunidad de Santa Mar³ya del Monte de Cea, Villamol y Villase	Le³n	Mancomunidad
12.170	Comunidad de Quintanaortu±o y Celadilla-Sotobr³yn (Merindad d	Burgos	Comunidad
12.171	Comunidad de Quintanaortu±o y Villanueva de R³yo Ubierna	Burgos	Comunidad
12.172	Comunidad de Quintanaortu±o, Sotopalacios y Villanueva	Burgos	Comunidad
12.173	Comunidad de Quintanilla-Vivar, Sotopalacios, Sotragero y Vi	Burgos	Comunidad
12.174	Jurisdicci³n Mancomunada de Bascu±ana y Viloría de Rioja	Burgos	Mancomunidad
12.175	Jurisdicci³n Mancomunada de Bascu±ana y Viloría de Rioja	Burgos	Mancomunidad
12.176	Jurisdicci³n Mancomunada de Bascu±ana y Viloría de Rioja	Burgos	Mancomunidad
12.177	Comunidad de Soto de la Vega y Villazala	Le³n	Comunidad
12.178	Comunidad de San Vicente del Valle y Villagalijo	Burgos	Comunidad
12.179	Comunidad de Riego de la Vega y Villamont³n de la Valduerna	Le³n	Comunidad
12.180	Mancomunidad de Las Quintanillas y Tardajos	Burgos	Mancomunidad
12.181	Comunidad de Rab³ de las Calzadas y Tardajos	Burgos	Comunidad
12.182	Comunidad de Torrelara y Villoruebo	Burgos	Comunidad
12.183	Comunidad de Quintanalara	Burgos	Comunidad
12.184	Comunidad de Cubillo del Campo y Hontoria de la Cantera	Burgos	Comunidad
12.185	Comunidad de Los Ausines, Quintanalara y Revilla del Campo	Burgos	Comunidad
12.186	Comunidad de Barbadillo del Pez y Jaramillo de la Fuente	Burgos	Comunidad
12.187	Comunidad de San Mill³n de Lara y Tinieblas de la Sierra	Burgos	Comunidad
12.188	Comunidad de Riocavado de la Sierra y Barbadillo del Pez	Burgos	Comunidad
12.189	Comunidad de Barbadillo de Herreros, Valderrubio y Villajime	Burgos	Comunidad
12.190	Comunidad de Jaramillo de la Fuente y Vizca³nos	Burgos	Comunidad
12.191	Ledan³ya de Huertas de Abajo y Tolba±os de Abajo	Burgos	Territorio Hist³rico

12.192	Comunidad de Hortig ³ ela, Jurisdicci ⁴ n de Lara y Mambril	Burgos	Comunidad
12.193	Comunidad de Barbadillo del Mercado, La Revilla y Pinila de	Burgos	Comunidad
12.194	Comunidad de Castrillo de la Reina, Hacinas y Salas de los I	Burgos	Comunidad
12.195	Comunidad de Castrillo de la Reina, Hacinas, Monasterio de I	Burgos	Comunidad
12.196	Comunidad de Barbadillo del Mercado, Cascajares de la Sierra	Burgos	Comunidad
12.197	Comunidad de Quintanilla del Agua	Burgos	Comunidad
12.198	Comunidad de Covarrubias y Quintanilla del Coco	Burgos	Comunidad
12.199	Comunidad de Barbadillo del Mercado y La Revilla	Burgos	Comunidad
12.200	Ledan ⁵ Ya de Hacinas, Salas de los Infantes y Villanueva de Carzo	Burgos	Territorio Histórico
12.201	Ledan ⁵ Ya de Barbadillo del Mercado, La Revilla, Salas de los Infa	Burgos	Territorio Histórico
12.202	Comunidad de Barbadillo del Mercado, Hacinas, La Revilla, Sa	Burgos	Comunidad
12.203	Comunidad de Quintanar de la Sierra y Vilviestre del Pinar	Burgos	Comunidad
12.204	Comunidad de La Revilla y Villanueva de Carazo	Burgos	Comunidad
12.205	Comunidad de Barbadillo del Mercado, La Revilla y Villanueva	Burgos	Comunidad
12.206	Comunidad de Carazo y Villanueva de Carazo	Burgos	Comunidad
12.207	Facer ⁶ Ya 24	Navarra	Territorio Histórico
12.208	Facer ⁶ Ya 191	Navarra	Territorio Histórico
12.209	Facer ⁶ Ya 42	Navarra	Territorio Histórico
12.210	Sierra de L ⁷ quiz	Navarra	Territorio Histórico
12.211	Facer ⁶ Ya 26	Navarra	Territorio Histórico
12.212	Facer ⁶ Ya 41	Navarra	Territorio Histórico
12.213	Facer ⁶ Ya 40	Navarra	Territorio Histórico
12.214	Facer ⁶ Ya 45	Navarra	Territorio Histórico
12.215	Facer ⁶ Ya 39	Navarra	Territorio Histórico
12.216	La Mancomunitat dels Quatre Pobles	Lleida	Mancomunidad
12.217	Facer ⁶ Ya 103	Navarra	Territorio Histórico
12.218	Facer ⁶ Ya 85	Navarra	Territorio Histórico
12.219	Facer ⁶ Ya 44	Navarra	Territorio Histórico
12.220	Facer ⁶ Ya 37	Navarra	Territorio Histórico
12.221	Facer ⁶ Ya 36	Navarra	Territorio Histórico
12.222	Facer ⁶ Ya 21	Navarra	Territorio Histórico
12.223	Facer ⁶ Ya 79	Navarra	Territorio Histórico
12.224	Facer ⁶ Ya 38	Navarra	Territorio Histórico
12.225	Facer ⁶ Ya 14	Navarra	Territorio Histórico
12.226	Facer ⁶ Ya 91	Navarra	Territorio Histórico
12.227	Facer ⁶ Ya 87	Navarra	Territorio Histórico
12.228	Facer ⁶ Ya 86	Navarra	Territorio Histórico
12.229	Facer ⁶ Ya 88	Navarra	Territorio Histórico
12.230	Comunidad de Amezketete y Ordizia	Guipó ⁸ zcoa	Comunidad
12.231	Sierra Aralar	Navarra	Territorio Histórico
12.232	Facer ⁶ Ya de Suarbe	Navarra	Territorio Histórico

12.233	FacerYá 16	Navarra	Territorio Histórico
12.234	FacerYá 9	Navarra	Territorio Histórico
12.235	FacerYá 10	Navarra	Territorio Histórico
12.236	FacerYá 2	Navarra	Territorio Histórico
12.237	Facero Aldape	Navarra	Territorio Histórico
12.238	FacerYá 63	Navarra	Territorio Histórico
12.239	FacerYá 52	Navarra	Territorio Histórico
12.240	FacerYá de Remendia	Navarra	Territorio Histórico
12.241	FacerYá Aristregui	Navarra	Territorio Histórico
12.242	Sierra Andia	Navarra	Territorio Histórico
12.243	Comunero de Ans ^¾ y Fago	Huesca	Territorio Histórico
12.244	Sierra Urbasa	Navarra	Territorio Histórico
12.245	FacerYá de Anezcar	Navarra	Territorio Histórico
12.246	FacerYá de Berriosuso	Navarra	Territorio Histórico
12.247	FacerYá 105	Navarra	Territorio Histórico
12.248	ParzonerYá de Encia	Álava	Territorio Histórico
12.249	Monte Com·n de las Amescoas	Navarra	Territorio Histórico
12.250	FacerYá 35	Navarra	Territorio Histórico
12.251	FacerYá 50	Navarra	Territorio Histórico
12.252	FacerYá 22	Navarra	Territorio Histórico
12.253	FacerYá 56	Navarra	Territorio Histórico
12.254	FacerYá 55	Navarra	Territorio Histórico
12.255	FacerYá 104	Navarra	Territorio Histórico
12.256	FacerYá 43	Navarra	Territorio Histórico
12.257	FacerYá 28	Navarra	Territorio Histórico
12.258	FacerYá 83	Navarra	Territorio Histórico
12.259	FacerYá 15	Navarra	Territorio Histórico
12.260	FacerYá 74	Navarra	Territorio Histórico
12.261	FacerYá 76	Navarra	Territorio Histórico
12.262	FacerYá 82	Navarra	Territorio Histórico
12.263	FacerYá 81	Navarra	Territorio Histórico
12.264	Comunidad de La±o	Álava	Comunidad
12.265	FacerYá 106	Navarra	Territorio Histórico
12.266	FacerYá 11	Navarra	Territorio Histórico
12.267	FacerYá 71	Navarra	Territorio Histórico
12.268	FacerYá 65	Navarra	Territorio Histórico
12.269	Facero Montejurra de Azqueta y Urbiola	Navarra	Territorio Histórico
12.270	FacerYá 70	Navarra	Territorio Histórico
12.271	Facero Arambelza	Navarra	Territorio Histórico
12.272	FacerYá 30	Navarra	Territorio Histórico
12.273	Comunidad de Barbarin y Olejua	Navarra	Comunidad

12.274	FacerYá 32	Navarra	Territorio Histórico
12.275	FacerYá de Lerga y UjuÚ	Navarra	Territorio Histórico
12.276	Bardenas Reales	Navarra	Territorio Histórico
12.277	Mancomunidad de Nalda, Sorzana y Viguera	La Rioja	Mancomunidad
12.278	Comunidad de la Pardina del Mercadal	Teruel	Comunidad
12.279	Jurisdicci4n de Segura de los Ba±os y Vivel del RYó MartYn	Teruel	Jurisdicci3n
12.280	Partzuergo Nagusia	Guip3zcoa	Parzonería
12.281	Partzuergo Txikia	Guip3zcoa	Parzonería
12.282	Comunidad de Castrillo de la Reina, Hacinas y Salas de los I	Burgos	Comunidad
12.283	Comunidad de Hacinas y Salas de los Infantes	Burgos	Comunidad
12.284	Comunidad de Canicosa de la Sierra, Quintanar de la Sierra y	Burgos	Comunidad
12.285	Comunidad de Hacinas y Pinilla de los Barruecos	Burgos	Comunidad
12.286	Comunidad de Castrillo de la Reina, Hacinas y Salas de los I	Burgos	Comunidad

Tabela 41 - Lista de Polígonos Isolados após Segmentação 1878-1940

Número do Polígono	Nome	Província / Distrito	País	Designação Administrativa	Cluster 1878-1940
225	Santa Cruz de Mudela	Ciudad Real	Espanha	Município	1
404	Chilluévar	Jaén	Espanha	Município	1
468	Larva	Jaén	Espanha	Município	1
888	Arriate	Málaga	Espanha	Município	1
1 806	Villatoya	Albacete	Espanha	Município	1
1 818	Cofrentes	Valencia	Espanha	Município	1
2 214	Armuña de Almanzora	Almería	Espanha	Município	1
2 726	Mortágua	Viseu	Portugal	Freguesia	1
2 785	Oliveira do Mondego	Coimbra	Portugal	Freguesia	1
3 242	La Fregeneda	Salamanca	Espanha	Município	1
3 785	Barcel	Bragança	Portugal	Freguesia	1
4 173	Pinhão	Vila Real	Portugal	Freguesia	1
6 498	Sotillo de las Palomas	Toledo	Espanha	Município	1
6 617	Villasequilla	Toledo	Espanha	Município	1
7 527	Muros de Nalón	Asturias	Espanha	Município	1
7 539	Avilés	Asturias	Espanha	Município	1
9 109	Berbinzana	Navarra	Espanha	Município	1
9 206	Arrúbal	La Rioja	Espanha	Município	1
9 307	Castellfollit de la Roca	Girona	Espanha	Município	1

9 312	Besalú	Girona	Espanha	Município	1
9 552	Gironella	Barcelona	Espanha	Município	1
9 596	Roda de Ter	Barcelona	Espanha	Município	1
10 125	Lumpiaque	Zaragoza	Espanha	Município	1
10 328	Pira	Tarrajona	Espanha	Município	1
10 463	Codo	Zaragoza	Espanha	Município	1
10 762	Lledó	Teruel	Espanha	Município	1
11 451	Benamaurel	Granada	Espanha	Município	1
924	Tolox	Málaga	Espanha	Município	2
2 287	Cordinhã	Coimbra	Portugal	Freguesia	2
2 308	Brenha	Coimbra	Portugal	Freguesia	2
2 309	Buarcos	Coimbra	Portugal	Freguesia	2
2 379	Memória	Leiria	Portugal	Freguesia	2
2 563	Eljas	Cáceres	Espanha	Município	2
2 590	Mesquitela	Guarda	Portugal	Freguesia	2
3 041	Amêndoa	Santarém	Portugal	Freguesia	2
3 911	Parada	Bragança	Portugal	Freguesia	2
6 919	A Illa de Arousa	Pontevedra	Espanha	Município	2
7 331	Nunes	Bragança	Portugal	Freguesia	2
7 685	Potes	Cantabria	Espanha	Município	2
8 987	Liédena	Navarra	Espanha	Município	2
9 046	Llivia	Girona	Espanha	Município	2
9 795	Aiguafreda	Barcelona	Espanha	Município	2
9 980	Sant Iscle de Vallalta	Barcelona	Espanha	Município	2
561	Zuheros	Córdoba	Espanha	Município	3
939	Igualeja	Málaga	Espanha	Município	3
2 246	Turrillas	Almería	Espanha	Município	3
3 856	Valverde	Bragança	Portugal	Freguesia	3
5 420	Garganta la Olla	Cáceres	Espanha	Município	3
6 131	Narros del Puerto	Ávila	Espanha	Município	3
6 326	Umbrías	Ávila	Espanha	Município	3
7 149	Santalha	Bragança	Portugal	Freguesia	3
7 759	Reinosa	Cantabria	Espanha	Município	3
7 763	Arija	Burgos	Espanha	Município	3
8 415	Barbadillo del Mercado	Burgos	Espanha	Município	3
8 470	Hacinas	Burgos	Espanha	Município	3
9 082	La Vajol	Girona	Espanha	Município	3
9 184	La Seu d'Urgell	Lleida	Espanha	Município	3
9 742	Castellnou de Bages	Barcelona	Espanha	Município	3
9 960	Aldealpozo	Soria	Espanha	Município	3
10 067	Muriel de la Fuente	Soria	Espanha	Município	3

465	Albánchez de Ubeda	Jaén	Espanha	Município	4
2 073	Paterna del Madera	Albacete	Espanha	Município	4
2 621	Videmonte	Guarda	Portugal	Freguesia	4
2 807	Fóios	Guarda	Portugal	Freguesia	4
3 693	Soeima	Bragança	Portugal	Freguesia	4
3 709	Burga	Bragança	Portugal	Freguesia	4
6 118	Collado del Mirón	Ávila	Espanha	Município	4
6 200	Casares de las Hurdes	Cáceres	Espanha	Município	4
6 278	Robledillo de Gata	Cáceres	Espanha	Município	4
6 334	Las Casillas	Ávila	Espanha	Município	4
7 114	Castro Laboreiro	Viana	Portugal	Freguesia	4
8 416	Monasterio de la Sierra	Burgos	Espanha	Município	4
8 753	Codos	Zaragoza	Espanha	Município	4
8 790	Zegama	Guipúzcoa	Espanha	Município	4
8 792	Leintz-Gatzaga	Guipúzcoa	Espanha	Município	4
9 027	Lapoblación	Navarra	Espanha	Município	4
9 128	Albanyà	Girona	Espanha	Município	4
9 197	Petilla de Aragón	Navarra	Espanha	Município	4
9 419	Bergasillas Bajera	La Rioja	Espanha	Município	4
9 741	Colluspina	Barcelona	Espanha	Município	4
9 764	San Felices	Soria	Espanha	Município	4
10 246	Quiñonería	Soria	Espanha	Município	4
10 320	Sestrica	Zaragoza	Espanha	Município	4
11 455	Alcóntar	Almería	Espanha	Município	4
11 932	Villanueva de Oscos	Asturias	Espanha	Município	4

8. BIBLIOGRAFIA

- Anselin, L. (1995). LOCAL INDICATORS OF SPATIAL ASSOCIATION - LISA. *Geographical Analysis*, 27(2), 93-115.
- Anselin, L., Bera, A. K., Florax, R., & Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26(1), 77-104. doi:10.1016/0166-0462(95)02111-6
- Ayuda, M. I., Collantes, F., & Pinilla, V. (2010a). From locational fundamentals to increasing returns: the spatial concentration of population in Spain, 1787-2000. *Journal of Geographical Systems*, 12(1), 25-50. doi:10.1007/s10109-009-0092-x
- Ayuda, M. I., Collantes, F., & Pinilla, V. (2010b). Long-run regional population disparities in Europe during modern economic growth: a case study of Spain. *Annals of Regional Science*, 44(2), 273-295. doi:10.1007/s00168-008-0260-9
- Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis*, 28(4), 281-298.
- da Silva, L. E., Alves, D., Lima, N. M., Alcantara, A., & Puig, J. (2011). Population and Railways in Portugal, 1801-1930. *Journal of Interdisciplinary History*, 42(1), 29-52.
- Da Silva, L. E., Alves, D., Painho, M., Costa, A. C., & Alcantara, A. (2013). The Evolution of Population Distribution on the Iberian Peninsula: A Transnational Approach (1877-2001). *Historical Methods*, 46(3), 157-174. doi:10.1080/01615440.2013.804787
- Fujita, M., & Mori, T. (1996). The role of ports in the making of major cities: Self-agglomeration and hub-effect. *Journal of Development Economics*, 49(1), 93-120. doi:10.1016/0304-3878(95)00054-2
- Han, J. (2006). *Data Mining*. USA: Elsevier.
- Krugman, P. (1991). INCREASING RETURNS AND ECONOMIC-GEOGRAPHY. *Journal of Political Economy*, 99(3), 483-499. doi:10.1086/261763
- Le Gallo, J., & Chasco, C. (2008). Spatial analysis of urban growth in Spain, 1900-2001. *Empirical Economics*, 34(1), 59-80. doi:10.1007/s00181-007-0150-5
- Marti-Henneberg, J. (2005). Empirical evidence of regional population concentration in Europe, 1870-2000. *Population Space and Place*, 11(4), 269-281. doi:10.1002/psp.373
- Miller, H. J. (2004). Tobler's First Law and spatial analysis. *Annals of the Association of American Geographers*, 94(2), 284-289. doi:10.1111/j.1467-8306.2004.09402005.x
- Moran, P. A. P. (1950). NOTES ON CONTINUOUS STOCHASTIC PHENOMENA. *Biometrika*, 37(1-2), 17-23. doi:10.2307/2332142
- Parracho, A. C., Melo Gonçalves, P., & Rocha, A. (2015). Regionalization of precipitation for the Iberian Peninsula: Elsevier.
- Pons, J., Paluzie, E., Silvestre, J., & Tirado, D. A. (2007). Testing the new economic geography: Migrations and industrial agglomerations in Spain. *Journal of Regional Science*, 47(2), 289-313. doi:10.1111/j.1467-9787.2007.00510.x
- Rappaport, J., & Sachs, J. D. (2003). The United States as a coastal nation. *Journal of Economic Growth*, 8(1), 5-46. doi:10.1023/a:1022870216673
- Ros, J. A., Rodríguez, P. C., Gisbert, F. J. G., & Mas, M. (2006). La localización de la población española sobre el territorio: un siglo de cambios: un estudio basado en series homogéneas (1900-2001).
- Ward, M., & Skrede Gleditsch, K. (2008). *Spatial Regression Models* (I. SAGE Publications Ed.).
- Wooldridge, J. (2009). *Introductory Econometrics*: South-Western Cengage Learning.